Diet, Emission and Diabetes: A treelet transform pattern analysis on Västerbotten Intervention Program

Author:

Wossenseged Birhane Jemberie

Supervisor/s:
Joacim Rocklöv, PhD (*Main Supervisor*)

Margareta Norberg, PhD (*Co-supervisor*)

Nr: 4/2018
Master thesis, 30 credits
Master's programme in Public Health, 120 credits

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2-h cPG: *2-hour post glucose load capillary plasma glucose*

AP: *Alcohol pattern*

BSP: *white bread and sweet spreads*

***cf*:** *carbon footprint*

cfAP: *carbon footprint from alcohol pattern*

cfFDB*:* *carbon footprint from fatty dairy & butter*

cfMP: *carbon footprint from mixed pattern*

cfPP: *carbon footprint from prudent pattern*

cfRFP: *carbon footprint from refined food pattern*

cfsum1: *greenhouse gase emissions in Kg CO_2e/day*

CO_2e: *carbon dioxide equivalent*

CV: *cross-validation*

DM: *diabetes mellitus*

FcPG: *fasting capillary plasma glucose*

ffoodtt: *dietary treelet component in women*

FFQ: *food frequency questionnaire*

GDM: *gestational diabetes*

GHGE: *greenhouse gas emissions*

HFdairy: *high fat dairy*

HFIbread: *high fiber bread and cereal*

IFG: *impaired fasting glucose*

IGT: *impaired glucose tolerance*

LCA: *life cycle assessment*

LFdairy: *low fat dairy*

LFIbread: *low fiber bread and cereal*

mfoodtt: *dietary treelet component in men*

MP: *mixed pattern*

OGTT: *oral glucose tolerance test*

PCA: *principal component analysis*

PP: *prudent pattern*

RFP: *refined food pattern*

T1DM: *type 1 diabetes*

T2DM: *type 2 diabetes*

TC: *treelet component*

TT: *treelet transform*

VIP: *Västerbotten Intervention Program*

Abstract

Objective: Researches which studied the relation of dietary greenhouse gas emissions with health outcomes are few, inconsistent and most of them are modelling studies which have not investigated empiric dietary emission patterns. In this study, we employ a posteriori data dimension reduction method, treelet transform, to identify dietary and diet related emission patterns concurrently. We aim to evaluate if these patterns are correlated, if they are associated with diabetes and if emission patterns can be used as a proxy for dietary patterns for assessment of association with diabetes.

Design: Food items from dietary questionnaire were aggregated to 34 food groups. GHGE was estimated by linking food intakes with life cycle assessment data on emission. Dietary and emission patterns were identified by employing treelet transform on food intake and corresponding greenhouse gas emission data. Multivariate logistic regression was performed to investigate associations between quintiles of dietary patterns and diabetes. Adjusted mean values of emission estimates were obtained for the identified dietary patterns. Adjusted proportions of diabetes across quintiles of emission patterns were computed.

Setting: Västerbotten Intervention Program

Subjects: women (*n* 38,118); men (*n* 36,042) between the age of 35 and 65 years

Results: Four dietary and four corresponding emission patterns in women, five dietary and five corresponding emission patterns in men were identified. Moderate to strong correlations were observed between dietary and corresponding emission patterns. Prudent dietary pattern (PP) in women was inversely associated with dysglycemia [$OR_{Q5\ vs.\ Q1}$ = 0.82 (95% CI 0.69—0.97, P_{trend} =0.003)]. PP in women was also inversely associated with diabetes [$OR_{Q5\ vs.\ Q1}$ = 0.37 (95% CI 0.17—0.78, P_{trend} = 0.002)]. However, adherence to this dietary pattern was associated with higher dietary emission. Finally, none of the corresponding emission patterns, were associated with adjusted proportions of either dysglycemia or diabetes.

Conclusion: Treelet transform produces correlated dietary and emission patterns which are sparse and easily interpretable. However, some differences in loading structures between dietary and emission patterns result in different conclusion regarding the association with diabetes, rendering the usage of emission patterns as proxies of dietary patterns inappropriate. Results from our study also show that healthy dietary patterns do not necessarily reduce greenhouse gas emission.

Keywords: Treelet transform; Dietary patterns; Greenhouse gas emissions; Carbon footprint; dietary GHGE patterns; Diabetes; Dysglycemia; Västerbotten Intervention Program

Introduction

Diabetes: Prevalence and Type

Diabetes mellitus (diabetes, DM) is a pandemic non-communicable disease, with nearly 500 million and increasing number of adults worldwide having the disease. In Sweden, the prevalence of diabetes is estimated to be 7% [1]. A recent research based on the study population in the Västerbotten Intervention Program (VIP) shows that the cumulative incidence proportion of diabetes was 3.3% at 10 years follow-up [2].

Diabetes occurs when blood glucose is higher than the normal level due to the inability of pancreas to produce any or enough insulin (Type 1 Diabetes, T1DM) or when the body is resistant to the produced insulin for maintaining blood glucose to normal level (Type 2 Diabetes, T2DM). Ultimately, in T2DM, also the insulin production fails. A third type of diabetes is gestational diabetes (GDM), an insulin resistance, which occurs during pregnancy. T2DM is the most prevalent form of diabetes, with 90% of diabetes patients being diagnosed as T2DM patients. For this reason, hereafter, we will continue to refer Type 2 Diabetes simply as diabetes or T2DM. Diabetes is a chronic disease which means it develops over long span of time. Most of the time, before the progression to diabetes, patients develop intermediate hyperglycemia, a prediabetic state. Prediabetes is diagnosed when blood glucose levels are higher than the normal level but below the threshold of diabetic blood glucose concentration. A higher than normal fasting blood glucose level indicates impaired fasting glucose (IFG) while impaired glucose tolerance (IGT) is diagnosed when a 2-hour postprandial blood glucose level is higher than it should normally be after ingestion of a 75-gram oral glucose load (This test is called oral glucose tolerance test, OGTT) [1]. In this study we refer the condition of IFG or IGT or T2DM as dysglycemia.

Diet: A risk factor for diabetes

Both diabetes and prediabetes share a range of risk factors. Accumulated evidence from prospective cohort studies and randomized controlled trials (RCTs) shows that one of the major modifiable risk factors is diet [1]. High intake of cereals rich in dietary fiber, vegetables and fruits have been shown to reduce the risk of diabetes while diets rich in refined and processed foods, red meat and added sugar are shown to increase diabetes risk. [3-8]. Early epidemiological researches studied the association of diet with diabetes by investigating single food or nutrient [7, 9-12]. However, despite its contribution to much of what we know about human nutrition's role in health status, this approach had its own methodological limitations. First, there is often collinearity between nutrients (for example, foods rich in fat often have low fiber content) which makes it difficult to conclude certainly that the studied food is in fact associated with diabetes. Second, the interactive effect of nutrients is likely to be greater than the sum of their individual effects, rendering the study of the whole diet necessary. Third, isolated nutrients are less bioavailable than the same nutrients in their natural dietary matrix. These resulted in the emerging of other research methods which complement the individual nutrient approach and focus on dietary patterns [3; 6].

The various dietary habits and combinations of consumed foods by population make dietary patterns complex. Their study has evolved from simple adequacy scores such as nutrient

adequacy ratio, NAR [13] to more complex diet scores, such as Healthy Eating Index, HEI [14], Dietary Approaches to stop Hypertension, DASH diet [15] and Mediterranean diet score, MDS [16,17] which all are evidence based *a priori* methods, i.e., they are developed to evaluate adherence of individuals to a recommended intake.

However, another reasoning argues that people choose what and how to eat and these population specific dietary patterns should be discovered from the data itself, a *posteriori* approach. Principal component analysis (PCA) and cluster analysis (CA) are the most widely used methods in empiric dietary pattern analysis. Recently, treelet transform (TT) [18], which is capable of extracting sparse and easily interpretable patterns based on the correlation (covariance) matrix of food components, has been introduced into nutrition epidemiology [19-21]. Though TT's suitableness to identify dietary patterns associated with disease outcomes has been shown in two studies, the only study which evaluated the association of TT driven dietary patterns with diabetes did not find any association, requiring further investigation of its applicability in another population context.

Many such *a posteriori* methods frequently identify patterns which are arbitrarily labeled as "western" or "unhealthy": often characterized by higher loading of red meats, refined foods and sugar, and another pattern labeled as "prudent" or "healthy": often characterized by higher loading of fruit, vegetables, high fiber cereals, fresh fish and poultry [6,22-24]. Often these similarly labeled patterns vary from study to study in the number of food components they are defined by or in the weight of the food components. Also, there is still inconsistency in results from these studies regarding disease association. However, generally, "western" dietary pattern has been associated with increased diabetes risk and "prudent" dietary pattern with lower diabetes risk [6,23,24].

Diet: Association with Greenhouse gas emissions

A quickly increasing global population and a shift in population dietary choice towards high-protein/high-fat diets mean more adverse effects on planetary health through land use change, biodiversity loss, water pollution and climate change. Despite the recognition of this threat, only handful of national dietary guidelines of high income countries include advises addressing the issues of sustainability [25, 26].

Due to availability of data through life cycle assessment (LCA), most of the researches on the environmental impact of food focus on diet related greenhouse gas emissions (GHGE) to the atmosphere. The significant diet related greenhouse gases include carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) [25]. The potential global warming impact of these gases, i.e., their global warming potential (GWP) is not the same. Methane has a GWP of 34 times that of CO_2 and nitrous oxide has a GWP of 298 times that of CO_2 [27]. As a result, their combined effect, which is often termed as carbon footprint (*cf*), is expressed as carbon dioxide equivalent (CO_2e), the amount of CO_2 which gives an equal GWP over a time span of 100 years [28].

Food consumption and food systems at large have tremendous environmental impact contributing up to 30% of global anthropogenic GHGEs [29]. Diets rich in animal protein and processed foods are reported to have higher carbon footprint than plant-based diets [30-33].

However, like in nutrition epidemiology studies, results from studies on diets with low GHGE and health outcome are not consistent. Moreover, studies linking GHGE and health outcome are very few compared to studies on dietary patterns.

A systematic review highlighting the importance to identify healthy diets with low GHGE found that 64% of measured health outcomes in included studies were worse for diets with low GHGE, suggesting that low GHGE diet scenarios may not be better than existing dietary patterns for disease prevention [34]. Another systematic review found out that while some of the reviewed studies on dietary carbon footprint reported a very modest reduction in mortality risk, there was no significant statistical association between low-carbon diets and health benefit [35]. However, another study [36] showed that adherence to World Health Organization (WHO) dietary recommendations would lead to a 17% reduction in GHGE. This study also stated that a drastic reduction in diet-related emissions is not realistic and requires major deviance from current population dietary patterns which renders such strategy unacceptable by the study population. A recent Swedish RCT trial on dietary intervention for weight loss among lactating women reported that weight loss regimen with calorie restriction did not result in dietary emission reduction [37] as the food change primarily targeted sugars which have low carbon footprint. The study also showed that the intervention which was in line with the 2004 Nordic Nutrition Recommendations resulted an increase in emission from fruits and vegetables in the intervention group.

Most of these researches which investigate the health co-benefits of low-carbon diets are modelling studies relying on population mean consumption values. They focus on individual food items, nutritional quality and extrapolate the health benefits from available literatures on nutritional guidelines, diets and health outcomes. It is only quite recently that studies started to link actual dietary choices among individuals with data from LCA [37-42]. Even so, these studies have not investigated whether dietary *cf* patterns are similar to dietary patterns in their component structure and if they can directly be used to estimate a disease risk or prevalence.

Objective of the study

The overall aim of our study is to apply an innovative method of data reduction, treelet transform (TT), for evaluation of associations between dietary and GHGE patterns on data from a large population based cohort, as well as their respective associations with dysglycemia and diabetes. By doing so, we will investigate the applicability of dietary emission patterns as a proxy to dietary patterns for assessment of association with diabetes.

Specifically, our research questions are the following:

1. Are TT-derived dietary and diet-related GHGE patterns among VIP participants correlated?
2. Is there any association between identified dietary patterns and prediabetes or diabetes?
3. Is there any association between identified GHGE patterns and prediabetes or diabetes?

Methodology

Study population

The study population included participants in the VIP [43]. VIP is an ongoing community-based health intervention program in Västerbotten county of Northern Sweden which was launched in 1985 -with the aim of reducing diabetes and cardiovascular diseases (CVD) related morbidity and mortality in the adult population. At ages 40, 50 and 60 years, all middle-aged inhabitants of the county are invited by their local health center for comprehensive health screening and counselling. Until 1995, individuals at the age of 30 years were also invited for the program. While health examination is conducted by health professionals at the center, participants complete extensive questionnaire which covers individual and environmental factors including lifestyle and food frequency questionnaire (FFQ) [43]. Between 1990 and 2016 a total of 172595 health examinations have been conducted within the VIP [44].

Exclusion Criteria

The self-administered, validated FFQ [45] initially included 84 food items, but since 1996 a shorter version with 64-66 food items has been used. At the time we obtained the dietary data (DietVIP; part of the VIP database) from the Northern Sweden Diet Database (NSDD) in 2015, only 29 observations were available for the year 2014. As a result, the current study population included observations between 1996 and 2013. During this time, a total of 95,214 observations (46,370 in men and 48,844 in women) with acceptable data on portion sizes, intake frequencies and food intake level (FIL) were available for our study. Details of the FFQ and the methodology followed to determine the acceptability of portion sizes and food intake level are explained elsewhere [46,47].

Afterwards, we merged individual biomedical, anthropometric, demographic and lifestyle variables from the VIP database with the nutritional variables DietVIP data. Due to the cross-sectional design of the current study, only unique observations (n=74,160; women= 38,118, men= 36,042) were used for the analysis, leading the exclusion of 17,840 observations by examining duplicated person unique identifiers due to participation twice or three times by the same individual at different ages. Moreover, individuals with: known diabetes mellitus diagnosis (n=1,080), fasting capillary plasma glucose level < 1 or > 25 mmol/L (n=19), 2-h capillary plasma glucose level < 1 or > 35 mmol/L (n=1,186), height < 130 or > 210 cm (n=21), body weight <35 Kg (n=7), waist circumference < 60 cm (n=23), total cholesterol < 0.5 or > 15 mmol/L (n=2), total serum triglycerides <0.15 or > 20 mmol/L (n=14), body mass index < 15 or > 70 Kg/m^2 (n=4) were excluded as proposed by the VIP guidelines. Finally, 795 participants below the age of 35 years and 63 individuals above the age of 65 years were excluded in alignment of the VIP objective.

Food Intake Measurement

Briefly, participants were asked to report how often they consumed different food and beverage items in average over the past 12 months with frequency of dietary intake provided in a 9-level scale ranging from “never” to “four or more times a day”. Dietary intake (g/day) of each participant for the food items was then calculated by multiplying the self-reported frequency by portion size values obtained from the Swedish food composition database [48] and by accounting self-reported portion sizes as indicated by photographs in

the FFQ. For this study, the 66 food items in the FFQ were further aggregated to 34 food groups according to nutrient content and cultural culinary choices [description of the food groups is shown in **Table 2**]. Dietary intake for each group [gram per day] was calculated by summing intake of original food items within the group.

Dietary GHGE estimation

The carbon footprint estimation for the 66 food items from the FFQ were calculated primarily based on the average Kg CO_2e values reported by Elin Röös [49] for various food items [Supplementary Table S1]. CO_2e values for food items which are imported to Sweden were calculated based on the detailed methodology as outlined in that report. For the food items whose *cf* estimates were not available in the report, we opted to use values from other reports published for Swedish setting [50,51]. When this was not possible (for the food items: pizza, cookies, pancakes), we performed calculations based on standard ingredient proportions obtained from the Swedish food composition database [48] and from a study in the UK [28] and based on CO_2e values of each ingredient as reported by Röös. Afterwards, we compared the results with emission values obtained from CleanMetrics™ food emission calculator—a database which has food production data both from Europe and the USA by excluding emission factors, which are not included in the Röös report, from life cycle stages [52]. A test of agreement between our GHGe data and that of Hoolohan's et.al study [28] resulted a Pearson's correlation of 0.91 and Lin's concordance correlation coefficient of agreement= 0.86 (95% CI= 0.8—0.92). This procedure was followed to minimize variations which arise from different life cycle assessment (LCA) studies as much as possible. The GHG emission values in our study are given in Kg CO_2e/Kg of food item. The equation which is used to calculate CO_2e values for each food item (j^{th} food item) among a VIP participant (i^{th} participant) is given as a product of the daily intake of the food item in kilograms and the GHGe value for a unit kilogram of that food item.

$$Daily\ Kg\ CO_2e_{(x_{ij})} = \left(\frac{Daily\ gram\ intake_{(x_{ij})}}{1000g}\right) * CO_2e_{(j)} \qquad (1)$$

However, for coffee and tea we have taken further considerations into account. While the FFQ used gram intake of liquid coffee and tea as unit of measurement, in the Röös report, GHGe values embodied in these food items were given per kilogram of dry weight (3Kg CO_2e/Kg dry weight). In order to calculate the emission values per kilogram of liquid, we followed the following steps.

1. We obtained unit portion size of liquid tea and coffee from the Swedish Food Composition Database (250g for tea and 150g for coffee).
2. We estimated the average amount of coffee powder used in a unit portion of liquid coffee as 7g (6g to 8g), based on a study by Doublet et al. [53]. For tea, we used the standard teabag sizes which are common in Swedish supermarkets (2g/teabag).
3. We calculated the GHGe values for the above portion sizes as 0.021 Kg CO_2e for 7g coffee powder (150g coffee drink) and 0.006 Kg CO_2e for 2g teabag (250g tea drink).

So, equation (1) was modified as:

$$Daily\ Kg\ CO_2e_{(x_{icoffee})} = \left(\frac{Daily\ gram\ intake_{(x_{icoffee})}}{150g}\right) * 0.021 \qquad (2)$$

And

$$Daily\ Kg\ CO_2e_{(x_{itea})} = \left(\frac{Daily\ gram\ intake_{(x_{itea})}}{250g}\right) * 0.006 \qquad (3)$$

Finally, the calculated dietary GHGe values for the food items were added accordingly to calculate the GHGe values of the aggregated food groups, hence new variables which represent these values were created for further analysis. The GHGe value for water was estimated zero as the main source of water in the study region is tap water which comes from fresh water resources. For this reason, a *cf estimates* dimension reduction analysis excludes water from the input variables (zero variance).

Ascertainment of Diabetes

Oral Glucose Tolerance Test (OGTT) with a 75-gram oral glucose load was performed on VIP participants with a fasting plasma glucose concentration (capillary) which is not indicative of diabetic state or who have not been previously known to have diabetes mellitus. Then, 2-hour capillary plasma glucose concentration of these participants was measured. Diabetes was assessed by using the WHO diagnosis criteria [54] with:

- Fasting capillary plasma glucose (FcPG) level ≥ 7.0 mmol/L ***or*** 2-h post glucose load capillary plasma glucose (2-h cPG) level ≥ 12.2 mmol/L indicating diabetes mellitus (DM)
- A FcPG level < 7.0 mmol/L ***and*** 8.9 mmol/L ≤ 2-h cPG < 12.2 mmol/L indicating Impaired Glucose Tolerance (IGT)
- A FcPG level ≥ 6.1 mmol/L and < 7.0 mmol/L, ***and*** 2-h cPG < 8.9 mmol/L indicating Impaired Fasting Glycemia (IFG).
- A FcPG level < 6.1 mmol/L ***and*** 2-h cPG < 8.9 mmol/L indicating normoglycemic state (healthy)

Afterwards, two dummy variables were created: one which combined participants with intermediate stages of hyperglycemia with diabetic individuals and another which consists of only diabetic and normoglycemic individuals.

$$\{DDM\_IFGorIGTasDM\} = \begin{cases} 0, & if\ normoglycemic \\ 1, & if\ IGT\ or\ IFG\ or\ DM\ (dysglycemic) \end{cases}$$

$$\{DDM\_withoutIFGorIGT\} = \begin{cases} 0,\ if\ normoglycemic \\ 1,\ if\ DM \end{cases}$$

Statistical analysis

Treelet Transform: Dimension reduction is a common and vital task during analysis as modern datasets are multidimensional and noisy while the order of variables does not have significant meaning attached to it. Dimension reduction is also important procedure when conducting research on dietary intake and health outcome. This technique results in derivation of dietary patterns from a dataset which contains multiple variables of food items and nutrients, assisting the investigation of the joint effect of dietary intake and the formulation of well-informed dietary guidelines.

A widely used dimension reduction method applied to dietary pattern extraction is principal component analysis (PCA). PCA provides successive best linear approximations by calculating the eigenvectors of the covariance or correlation matrix of the observed variables and by producing components onto which the data is projected (i.e., linear combinations of the variables with maximal variance with the last component having the lowest variance of all linear combinations of the variables). However, the involvement of all variables in all the components (i.e., lack of sparsity) makes interpretation of components and detection of internal localized structures in the data difficult. While cluster analysis could be helpful in identifying hidden variable groupings, it lacks the ability to generate summary variables which quantifies the patterns. Moreover, in cluster analysis, patterns are extracted from individuals' intake difference instead of the difference between the observed variables as in case of PCA [*55*] thus addressing different research question.

Treelet transform (TT) is a relatively recent technique of dimension reduction which is proposed by Lee et al [*17*]. TT generates components by aggregating observed variables according to correlation. Though treelet components (TCs) are ordered based on the variation they explain, they differ from principal components (PCs) as they are sparsely loaded, simplifying interpretation. In TT, the numeric value of a variable within a component is called the variable loading. The TT algorithm simplifies interpretability by also producing associated binary cluster tree which visualizes the hierarchical joining of variables.

Briefly, TT algorithm works by identifying two original food items with the highest correlation (or covariance) and performing local PCA on them. Then the two variables are merged and replaced with low resolution "sum variable" and high resolution "difference variable". The sum variable is retained while the difference variable (residual) is not processed. This procedure is iterated until all the variables join the hierarchical cluster tree. TT differs from PCA as it does not automatically generate components with higher variance. Components are rather extracted by selecting a cluster tree level (choosing cut-level). Cutting the cluster tree near to the leaves produces factors with higher sparsity but less explained variation. Meanwhile, choosing a cut-level near to the roots results in densely loaded components (lower sparsity) rendering interpretation difficult. Hence, the use of cross-validation (CV), a data-driven choice of cut-level. Lee et al. [*17*] propose first to cut the tree at a given level and extract components at this level based on their variance. Afterwards, a K-fold cross-validation technique, in which the data is randomly split into K roughly equal-sized subsets, is applied. Then, for each possible cut level ([1, *n*-1]; where n is number of variables) and "*m*" number of retained components, the *m* highest variance components are calculated by using data from

K-1 subsets. Next, the sum of variances of scores based on these components are calculated using only the remaining subset. This procedure is repeated K times, each time omitting a different subset. Subsequently, the CV score at each possible cut level is calculated by averaging the resulting K sums of component variances. Once CV scores have been calculated, an optimal cut-level can be obtained from a graph of CV scores vs cut-level by locating the point ("knee") where an increase in cut-level does not yield a substantial increase in explained variation by the retained components.

Dietary and GHGE Pattern Identification

The Treelet Transform (TT), a linear dimension reduction technique, was used in the current study to extract both dietary and carbon footprint (*cf*) patterns for men and women participants separately.

By applying TT to the correlation matrix of the dietary intake values across the 34 food groups and their respective carbon footprint estimates, we produced orthogonal components which could reveal the intrinsic structure of the observations. The sparsity of these components and the visualization of the grouping structure among the food items (and their respective cf) with the help of cluster tree simplify interpretability. However, it is important to note that unlike PC scores, TT scores are correlated, albeit weakly. Percentage of explained variance, Scree plot of variances and interpretability of components were the criteria used to extract and retain dietary and *cf* patterns. As our sample size was big, the optimal cut level for TT was calculated by using a 10-fold CV technique (to reduce bias and overfitting) and retaining CV scores for that cut level which are within 5% of the CV score for the n-1 (the maximal) cut level. Furthermore, to reduce inherent sampling variation without increasing bias, we performed 10 Monte-Carlo repetitions of cross-validation.

Sensitivity analysis and stability assessment: We performed sensitivity analysis by repeating TT at ±3 of the optimal cut level to investigate the effect of selecting different cut levels on component loadings. Finally, since sparsity is the key feature of TT, we proceeded to assess the stability of the obtained components by subsampling. First, we calculated the sign patterns of the obtained loadings for each component and the variance explained by that component. Second, TT was performed on a random sample of 80% of the original data. This procedure was repeated in 500 subsamples. The number of times the original sign patterns appear among all 500 groups is used as a measure of stability.

Associations between type 2 Diabetes vs dietary and GHGE patterns

After extracting the patterns, component scores were categorized into quintiles, with individuals in the highest quintile (Q5) showing greater adherence to that particular dietary pattern than those in the lowest quintile (Q1). The association between diabetes mellitus and retained dietary patterns was investigated by using multivariate logistic regression by adjusting for potential confounders: total energy intake in Kcal/day including from alcohol (continuous), physical activity (Cambridge index:1=Inactive[reference] 2=Moderately inactive 3=Moderately active 4=Active), age group (35-44 years[reference], 45-54 years, 55-64 years), education level (≤9 years[reference], upper secondary education, higher education), smoking status in cigarette/day (Nonsmoker[reference], 1-4, 5-14, 15-24, ≥25, occasional smoker,

former occasional smoker, former smoker), alcohol intake in g/day(≤20[reference], abstain/>20), fiber intake in g/1000Kcal (<15, ≥15[reference]), component scores of other dietary patterns (continuous); potential effect modifiers: body mass index in Kg/m2 (underweight: BMI <18.5; normal[reference]: 18.5 ≤ BMI <25; overweight: 25 ≤ BMI <30; obese: BMI ≥30), medication for hypertension (yes, no[reference), medication for dyslipidemia (yes, no[reference]) in the final model. Events per variable (EPV) for each covariate were assessed to avoid sparse data bias. Separate analyses were conducted for the two diabetes outcomes (Dysglycemia =1; DM=1). Tests for trend were computed by inputting quintile levels of component scores to the final model as a continuous variable, and statistical significance was set to *α*=0.05. The TT add-on for Stata, developed by Gorst-Rasmussen [56] was used to perform treelet transform on the food and carbon footprint variables.

The relationship between total dietary carbon footprint (*cf*) and dietary component scores was investigated by using the user-written **predxcat** [57] Stata add-on. The mean value of *cf* across quintiles of dietary component scores was obtained before and after adjusting for baseline covariates whose crude distribution across quintiles and significance levels are shown in Tables 5, 6 and supplementary tables S6 and S7. Statistical significance for the test for difference of 5 means in the adjusted analysis was investigated by using 4-degrees of freedom partial F-test for an overall association between *cf* estimate and quintiles of dietary component scores. Finally, we obtained the crude and adjusted proportions of diabetes outcome at quintiles of *cf* component scores. The mean values of covariates were used in the regression models during the adjustment process. The p-value for the adjusted analysis was computed from 4-degrees of freedom likelihood ratio test for difference of 5 proportions. All statistical analyses were conducted using Stata version 13.1 [58].

The proposed approach in our study is shown in **Figure 1**.

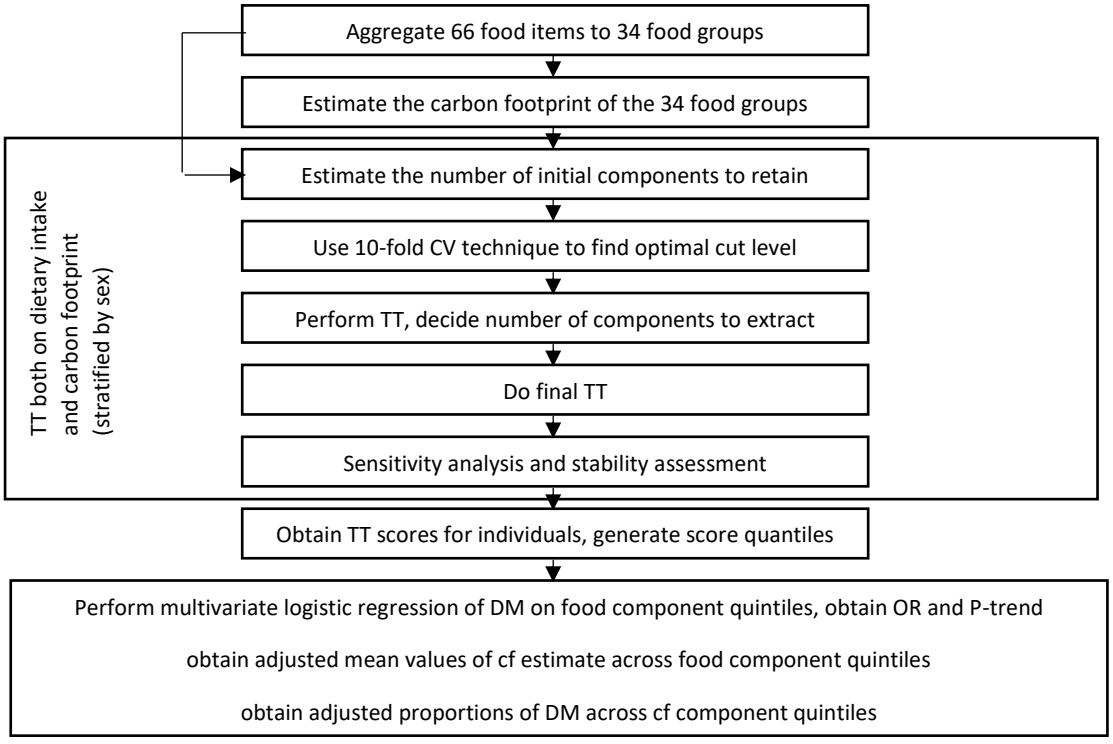


Figure 1. Flowchart depicting the statistical methods and steps used in the study

Results

Dietary and carbon footprint patterns

Information on basic characteristics of VIP participants is shown in **Table 1**. 50.5% of women had normal BMI while 49.4% of men were overweight. The proportions of women and men who had a fiber intake of more than 15 gram per 1000 Kcal were 20.8% and 5% respectively. Moreover, more women, compared to men, were with higher education (37.4% vs 26.7%) and current smokers (14.3% vs 11.2%).

Table 1: Characteristics of participants in the Västerbotten Intervention Program (VIP) between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables

Characteristic	**Women** *(n= 38,118)*	**Men** *(n=36,042)*
Height (cm)	165.1 ± 6.1	178.7 ± 6.6
Weight (kg)	70.1 ± 12.9	85.2 ± 13.1
BMI group (%)		
underweight	1.1	0.2
normal	50.5	34.6
overweight	32.8	49.4
obese	15.6	15.7
Reported energy intake (Kcal/day)	1,492.9 ± 393.6	1,992.6 ± 547.0
Estimated GHGe (Kg CO_2e/day) †	2.7 ± 0.9	3.4 ± 1.2
Alcohol intake (0 or >20g/d) (%)	9.4	8.3
Fiber intake (<15g/1000Kcal) (%)	79.2	95.0
Macronutrient intake (% of total energy intake)		
Protein	15.3 ± 2.3	14.5 ± 2.3
Carbohydrate	49.2 ± 7.1	45.4 ± 7.3
Fat	33.3 ± 6.6	36.8 ± 7.0
Alcohol	1..5 ± 1.7	2.3 ± 2.2
Age group (%)		
35-44 years	35.4	34.7
45-54 years	31.8	31.2
55-64 years	32.9	34.1
Education level (%)		
≤9 years	30.1	36.3
high school	32.5	37.0
higher education	37.4	26.7
Physical activity (%)		
inactive	16.9	17.6
moderately inactive	29.0	28.5
moderately active	27.1	29.3
active	26.9	24.6
Smoking status (%)		
non-smoker	50.9	53.0
1-4 cigarette	2.7	2.6
5-14 cigarette	9.1	5.2
15-24 cigarette	2.4	3.1
25+ cigarette	0.1	0.3
occasional smoker	2.7	3.3
former occasional smoker	10.4	9.9
former smoker	21.7	22.7
Medication for hypertension (%)	14.0	14.4
Medication for dyslipidemia (%)	3.1	5.8
Dysglycemia (%)	18.2	20.1
Diabetes (%)	1.5	1.9

† *the estimated mean GHGe for participants of both sexes in the VIP was 3.02 ± 1.1.*

For further analysis, the 66 food items in the FFQ were re-grouped into 34 food items. Description of the aggregated food items and the median intake values for men and women is given in **Table 2**. Important differences in median daily dietary intake values between men and women included high fat dairy products [96 g vs 45 g]; low fat dairy products [201.5 g vs 178.6 g]; butter [12.9 g vs 5.02 g]; low fiber cereals [17.5 g vs 9.7 g]; fruits [95.4 g vs 157.1 g]; vegetables [51.06 g vs 107.1 g]; beer [68 g vs 2.67 g]; soda [37.33 g vs 18 g]; tea [35 g vs 90g] and water [625 g vs 1000g].

Table 2: Description of food groups included in the study and median values of daily dietary intake with 25th and 75th percentiles in parentheses

Variable	Food Items	Intake [g/day]	
		Men	Women
HFdairy	Dairy: Sour milk, yoghurt, milk, 3% fat	96 (22.4, 229.2)	45 (17.6, 177.6)
LFdairy	Dairy: Sour milk, yoghurt, milk, low fat	201.5 (81, 475)	178.57 (89.1, 385)
butter	Bregott & Butter on bread; butter for cooking	12.9 (2.06, 26.6)	5.02 (0.09, 16.84)
margarine	margarine on bread, margarine for cooking	11.6 (2.83, 24.46)	7.04 (1.4, 17)
oil	Oil for cooking, salad dressing with oil	6.96 (2.8, 14.07)	5.28 (2.16, 10.28)
dressing	Cream, Crème fraiche, sour cream	3.2 (2.52, 6.48)	3.5 (2, 9)
HFIbread	Whole grain crisp and soft bread, whole wheat, oat, rye or barley porridge, fiber cereals	91.8 (55.98, 152.5)	88.2 (55.57, 146)
LFIbread	White crisp and soft bread, corn flakes	17.46 (6.23, 33.88)	9.7 (3.43, 22.04)
cheese	Cheese (medium/high fat)	14.06 (6.89, 20.06)	12.6 (5.76, 16.06)
bakery	Coffee rolls, rusk, cookies, pastry	11.62 (6.64, 29.16)	8.5 (5.52, 19.36)
processedMeat	Sausage, liver pate, meat on bread, bacon	25.8 (16.99, 38.75)	17.36 (11.46, 25)
soups	Rosehip, sweet syrup soup	0.6 (0.53, 14)	0.41 (0, 9.2)
fruits	Berries (fresh/frozen), apple, pear, orange, banana	95.4 (49.99, 159.2)	157.1 (99.5, 303.5)
vegetable	Root, fruity, leafy, stalk vegetables and cabbages	51.06 (18.7, 101.6)	107.1 (57.04, 190.8)
potato	Potatoes	105.2(64.6, 172.3)	97.1 (57.9, 160.3)
pasta	Pasta and rice	64.6 (37.1, 108.7)	56.2 (30.4, 95.4)
classic	Legumes and Swedish dumpling	31 (21.74, 40.2)	22.85 (12.83, 37.93)
fastfood	Pizza and hamburger	34.8 (10.58, 35.6)	12.33 (0.99, 21)
redmeat	Minced meat, meat stew, steak	52.5 (36, 80)	38.5 (26.5, 60.5)
poultry	Poultry	15.4 (12, 21)	15.4 (8.8, 21)
icecream	Ice-cream	5.44 (5.12, 5.76)	5.2 (0.2, 5.2)
sweets	Sweets (chocolate, candy)	3.5 (2, 7.2)	3.08 (1.76, 7.92)
sugar	Sugar, honey, marmalade, jam	2.73 (1.56, 18)	2.03 (1.16, 10.15)
saltysnack	Chips, popcorn, salted nuts	2.8 (0.11, 4.9)	2.16 (0.08, 3.78)
soda	Sodas	37.33 (21.33, 96)	18 (15.3, 69)
coffee	Brewed and boiled coffee	375.45 (375, 600)	375.45 (204, 480)
tea	Tea	35 (0.75, 175)	90 (20, 250)
beer	All types of beer	68 (29.5, 120.99)	2.67 (0.6, 35.15)
wine	Red and white wine	22 (0.83, 22)	18 (0.68, 18)
liquor	Liquor and spirits	0.25 (0.25, 6.72)	0.15 (0, 0.15)
water	Water	625 (625, 1000)	1000 (625, 1000)
egg	Eggs	7 (4, 7)	7 (7, 18)
freshfish	Lean and fatty fish	20.2 (11.56, 28)	17.2 (10.02, 23.8)
processedFish	Salted and smoked fish	0.38 (0.2, 3.68)	0.26 (0.2, 2.18)

After performing TT separately for male and female participants, 4 dietary patterns and 4 *cf* patterns, explaining 21.42% and 21.52% of the variance respectively, were identified among female participants while 5 dietary patterns and 5 *cf* patterns, explaining 24.96% and 25.37% of the variance respectively, were identified among male participants (percentage component variances for women and men are shown in **Table 3** below). TT on *cf* estimates excluded water from the analysis because of zero variance. These patterns were retained based on their variance (variance greater than 1); and after performing cross-validation, sensitivity analysis at ±3 of the optimal cut-levels obtained from CV and stability assessment of components. This procedure along with observation of scree plots facilitated interpretability of the retained patterns.

Table 3: Percentage Component Variances for TT on Dietary and GHGe data from female and male participants in the VIP

		Components					
		TC1	TC2	TC3	TC4	TC5	Total
*Dietary Patterns**	Women	7.07	5.48	4.93	3.94	-	21.42
	Men	7.71	5.36	4.30	4.15	3.44	24.96
*Carbon footprint patterns**	Women	7.26	6.08	4.12	4.06	-	21.52
	Men	7.85	5.69	4.43	3.99	3.41	25.37

TC1: Treelet Component 1; TC2: Treelet Component 2; TC3: Treelet Component 3; TC4: Treelet Component 4; TC5: Treelet Component 5
* proportions are adjusted for correlation between component scores using Gervini and Rousson's method (2004)

Results obtained from cross-validations indicated that the optimal cut-levels for dietary cluster tree were 24 for women and 22 for men while the optimal cut-levels for the *cf* estimates cluster tree were 25 for women and 23 for men [**Figure 2**].

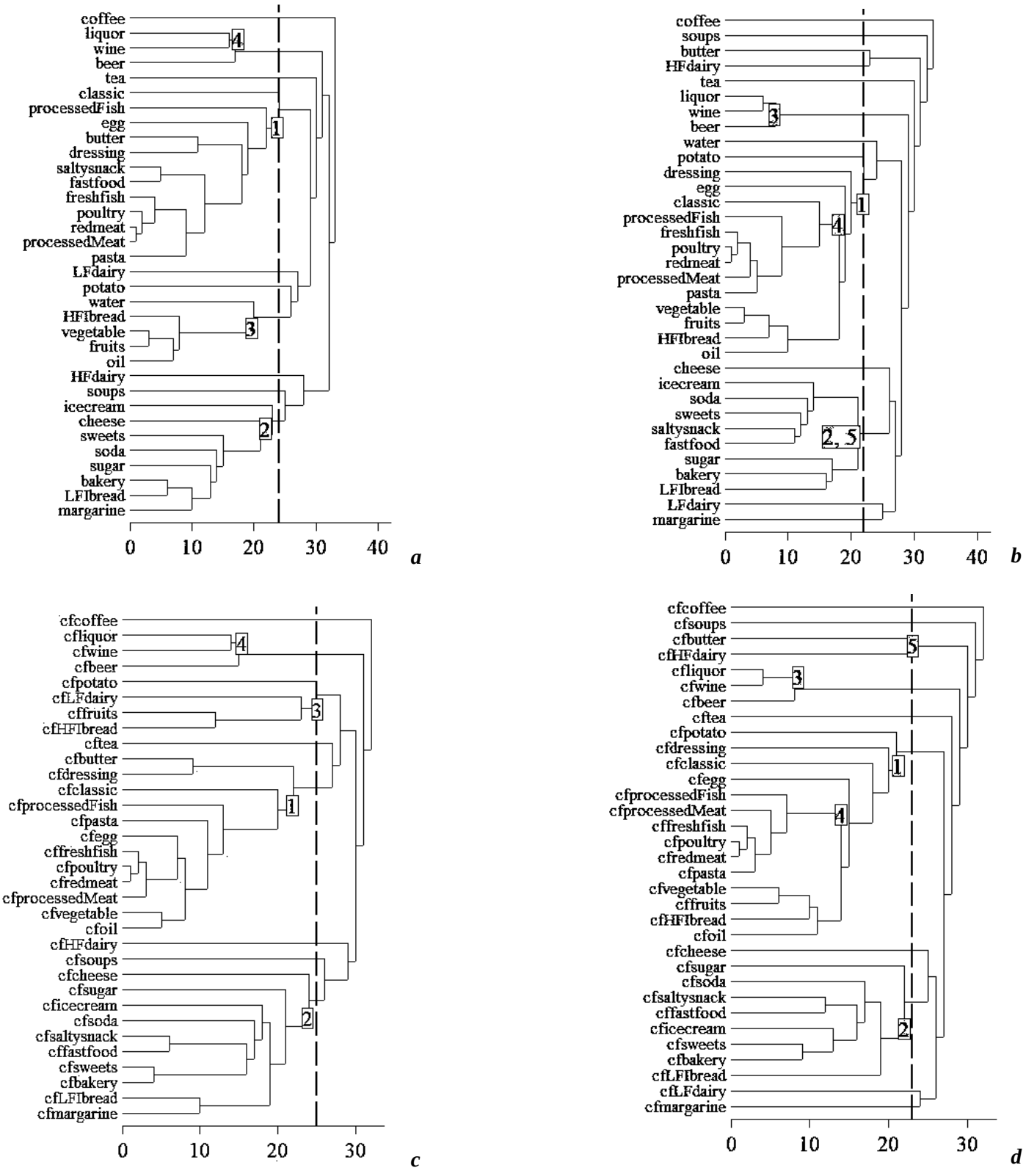
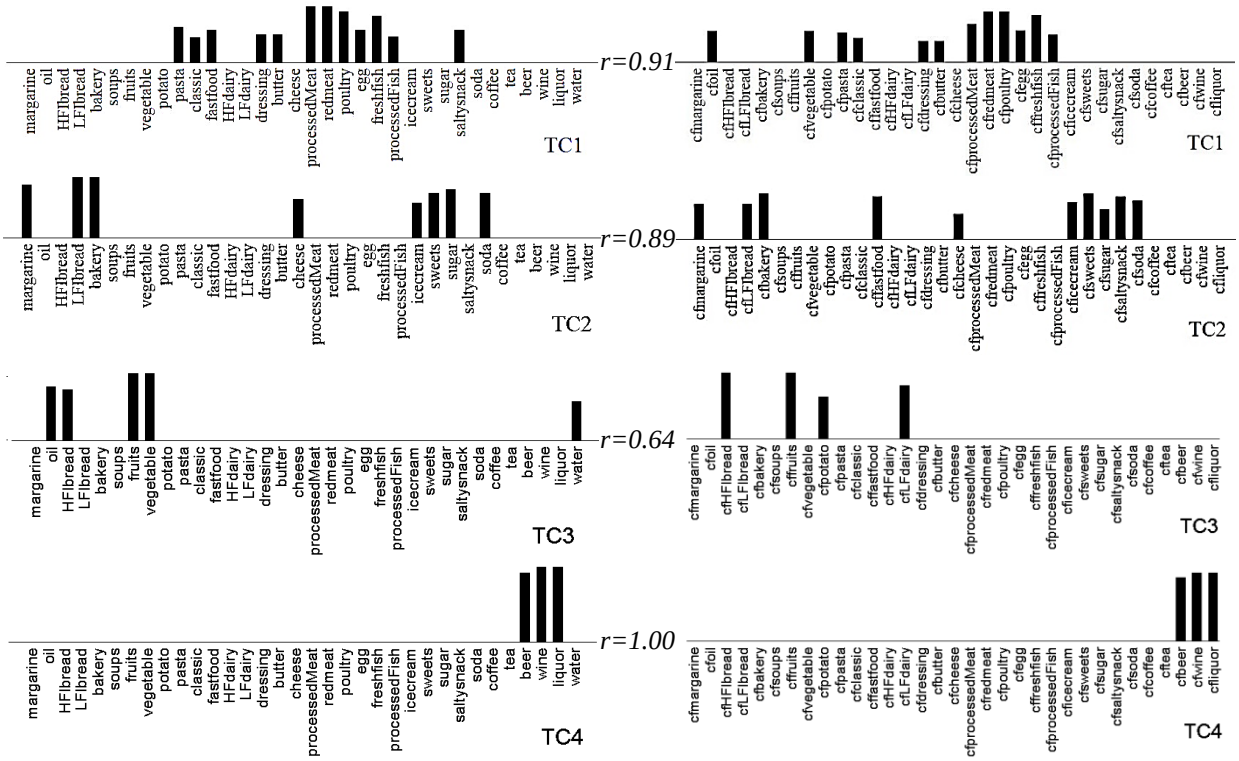


Figure 2. Cluster trees produced by the TT algorithm applied to dietary (a &b) and cf estimates (c & d) data in female (a & c) and male (b & d) participants in the VIP between 1996 and 2013. The vertical dashed lines indicate the selected cut levels for the cluster trees at which the components with the highest variance (indicated in the numbered nodes) are extracted. Leaves related to these nodes have non-zero loading on the given component.

Female: dietary (left) and cf patterns (right)



Male: dietary (left) and cf patterns (right)

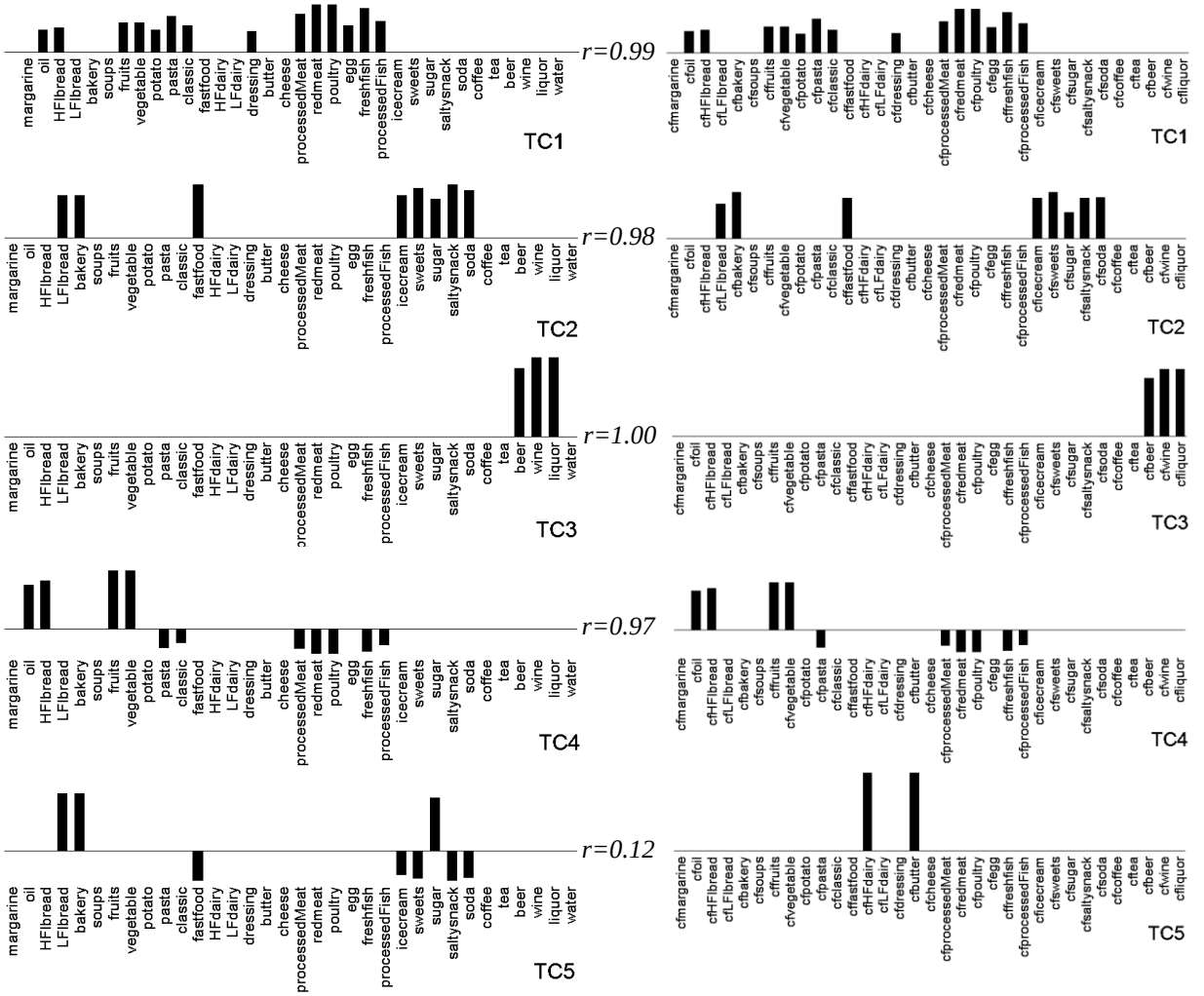


Figure 3. Treelet Component(TC) loading plots for dietary and cf patterns among male and female participants in VIP, obtained from TT. The horizontal axis is oriented such that loadings above the line are positive while those below the line are negative.

[**Figure 3**] The first dietary treelet component (TC1) among female participants was characterized by high intake of animal protein and fat (processed meat, hamburger, pizza, red meat, poultry, fish, egg, butter and crème), pasta and rice as carbohydrate source, classic foods (dumplings, waffles and baked beans) and salty snacks. The structure of the first *cf* component was characterized by high *cf* estimates of those food items which loaded on dietary TC1 except the *cf* estimate from fast foods and salty snacks which had zero loading. Instead, *cf* estimates from oil and vegetables had high loading. The correlation between dietary and *cf* TC1 scores was 0.91. Among male participants, the first dietary component also included oil, high fiber cereal products, fruits, vegetables and potatoes in addition to food items observed in dietary TC1 among female participants, while butter, fast foods and salty snacks were excluded from TC1 for men. The correlation between dietary and *cf* TC1 among men was 0.99. Because of their structure the dietary patterns are labelled "*Mixed Pattern,* ***MP***" and the cf patterns are labelled cf from Mixed dietary pattern, "***cf*MP**"

The second dietary and *cf* treelet components (TC2) among women were characterized by high intake of refined foods, cheese and sugar, and the *cf* resulting from the consumption of these foods respectively. *Cf* TC2 included emission from fast foods and salty snacks as well. The correlation between the scores of dietary and *cf* TC2 was 0.89. For men, the second dietary and *cf* components (TC2, correlation= 0.98) had similar characterization with those observed in women, however the dietary component did not load on margarine and cheese but included fast foods and salty snacks; the *cf* pattern was however sparser as emissions from margarine and cheese were omitted. Accordingly, these patterns were labelled "*Refined Food Pattern,* ***RFP***" and cf from "*Refined Food Pattern,* ***cf*RFP**"

A distinct TC characterized by exclusively high intake of alcohol (Alcohol pattern, **AP**) was obtained for women (TC4) and men (TC3). Moreover, these components had a correlation-between-scores of 1.00 with the counterpart TCs for *cf* estimates (***cf*AP**) for the respective sexes.

The dietary and *cf* TC3 among women were moderately correlated (r=0.64) as the dietary TC3 was characterized by high intake of vegetable oil, high fiber cereals, fruits, vegetables and water, hence labelled "*Prudent Pattern*, **PP**". The *cf* TC3 (***cf*PP**) had non-zero loading on high fiber cereals, fruits, potatoes and low-fat dairy products. For men, TC4 had zero loading for water, while higher consumption of meat and fish products led to reduced component score (negative loading). This component, labelled again **PP** which is characterized by higher intake of vegetable oil, high fiber cereals, fruits and vegetables, had a correlation of 0.97 with the corresponding *cf* component, ***cf*PP**.

Finally, the fifth dietary component (TC5) among men was characterized by positive loading of low fiber bread, bakeries, jam, marmalade, honey and sugar (hence labelled as *"white bread and sweet spreads"*, **BSP**), and negative loadings of fast-foods, ice cream, sweets, salty snacks and soda— a structure which makes its characterization as certain dietary pattern difficult. It had a very weak correlation (r=0.12) with the fifth *cf* component which loaded only on *cf* values of high fat dairy & butter (cf from fatty dairy & butter, ***cf*FDB)**.

Results from stability and sensitivity analysis

Stability test [**Table 4**] for the retained components showed that most of the retained components had appeared in more than 70% of the subsampling repetitions except the first and second dietary patterns among women. These patterns (frequency= 64.4% and 64.8% respectively) competed with two other patterns: one which removed fast food and salty snacks from the first pattern (frequency= 33.4%) and another which loaded the two food groups to the second pattern which appeared in 33.4% of the subsampling repetitions. These competing patterns were not retained. A sensitivity analysis at cut level 20 (23±3) showed that the fifth *cf* component in men was replaced by another competing component indicating its instability at cut level 20 despite its appearance in 94.2% of the subsampling repetitions at the optimal cut level. A sensitivity analysis at ±3 of the selected cut levels for other components did not result any other instability except that the discussed components became sparser (for -3 cut-level) or denser (+3 cut-level). A slightly different component ordering was observed for the *cf* patterns among women when a sensitivity analysis was performed at cut-level 22.

Table 4: Results from Stability analysis for retained Dietary and GHGe Treelet Components for female and male participants of VIP

		Components				
		TC1	TC2	TC3	TC4	TC5
Dietary Patterns[*]	Women	64.4	64.8	99.2	100	-
	Men	88	98.2	100	81	98
Carbon footprint patterns[*]	Women	93.6	95	94.6	100	-
	Men	96.4	98	100	76.2	94.2

TC1: Treelet Component 1; TC2: Treelet Component 2; TC3: Treelet Component 3; TC4: Treelet Component 4; TC5: Treelet Component 5
*Frequencies of retained component sign patterns among subsampled component sign patterns

Distribution of characteristics of participants across quintiles of dietary patterns

The unadjusted distribution of the characteristics of the study participants across quintiles of dietary patterns for women and men is described in **Table 5** and **Table 6**. In both women and men, individuals with a higher score for MP were found to be younger, more educated, physically active, never-smokers, have higher BMI, lower alcohol and fiber intake. The proportions of medication for hypertension, medication for dyslipidemia, dysglycemia and diabetes were lower in individuals who had higher score for dietary MP.

For RFP, in both women and men, individuals at the highest quintile were younger, never-smokers but physically inactive, had higher alcohol but lower fiber intake. The proportions of medication for hypertension and medication for dyslipidemia were lower for these individuals. Differences in distribution of characteristics were also found between women and men. The highest quintile for RFP in women consisted of more obese and less educated individuals while in men it was characterized by lower proportion of obese and higher proportion of more educated. Furthermore, in men, the highest quintile for this pattern had higher proportion of individuals with both dysglycemia and diabetes while in women there was no statistically significant difference in distribution across quintiles of scores.

***Table 5**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of dietary patterns in female participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables*

	"Mixed pattern" in women				"Refined food pattern" in women			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	1.88 ± 0.39	2.54 ± 0.47	3.79 ± 1.08	<0.001	2.50 ± 0.91	2.60 ± 0.85	2.98 ± 0.98	<0.001
Energy intake (Kcal/d)	1238.74 ± 256.44	1453.83 ± 312.33	1851.16 ± 432.14	<0.001	1272.79 ± 295.07	1445.04 ± 331.08	1818.30 ± 415.90	<0.001
Age group				<0.001				<0.001
35-44 years	14.4%	37.2%	52.7%		26.5%	35.2%	43.6%	
45-54 years	31.7%	32.1%	29.4%		31.0%	32.4%	33.2%	
55-64 years	53.9%	30.7%	17.9%		42.5%	32.4%	23.2%	
BMI group				<0.001				<0.001
underweight	1.2%	1.1%	1.2%		0.8%	0.9%	1.7%	
normal	52.1%	51.5%	47.8%		49.9%	51.3%	51.1%	
overweight	34.8%	32.4%	30.9%		34.7%	32.6%	30.1%	
obese	11.9%	14.9%	20.1%		14.6%	15.3%	17.1%	
Alcohol intake (0 or >20g/d)	12.1%	8.2%	9.4%	<0.001	6.9%	8.2%	13.6%	<0.001
Fiber intake (<15g/1000Kcal)	62.1%	81.4%	90.5%	<0.001	61.5%	80.4%	94.6%	<0.001
Education level				<0.001				<0.001
≤9 years	45.4%	28.8%	18.1%		27.9%	28.9%	32.8%	
high school	23.8%	33.9%	37.7%		30.8%	33.1%	34.7%	
higher education	30.8%	37.3%	44.2%		41.3%	38.0%	32.5%	
Physical activity				<0.001				<0.001
inactive	18.8%	16.8%	14.8%		15.5%	17.1%	18.3%	
moderately inactive	32.4%	29.4%	25.7%		27.0%	29.0%	32.0%	
moderately active	25.9%	27.6%	27.2%		25.7%	27.0%	27.7%	
active	22.8%	26.3%	32.3%		31.8%	26.9%	22.0%	
Smoking status				<0.001				<0.001
non-smoker	48.2%	50.5%	54.6%		45.2%	51.2%	54.2%	
1-4 cigarette	3.1%	2.7%	2.2%		2.7%	2.7%	2.6%	
5-14 cigarette	11.2%	8.9%	6.8%		8.7%	8.8%	10.1%	
15-24 cigarette	2.4%	2.6%	2.2%		2.0%	2.2%	3.5%	
25+ cigarette	0.1%	<1%	0.1%		0.1%	0.1%	0.1%	
occasional smoker	2.3%	2.6%	2.8%		2.9%	2.7%	2.5%	
former occasional smoker	10.2%	10.2%	10.9%		13.0%	10.9%	7.8%	
former smoker	22.6%	22.4%	20.4%		25.4%	21.4%	19.3%	
Medication for hypertension	16.9%	14.0%	11.4%	<0.001	16.6%	14.1%	10.8%	<0.001
Medication for dyslipidemia	3.7%	3.0%	2.5%	<0.001	4.4%	3.0%	1.5%	<0.001
IFG or IGT or Diabetes	19.5%	17.7%	17.3%	<0.001	18.9%	17.7%	17.9%	0.28
Diabetes	1.7%	1.4%	1.2%	0.004	1.5%	1.3%	1.4%	0.71

***Table 5**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of dietary patterns in female participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables (**continued**)*

	"Prudent pattern" in women				"Alcohol pattern" in women			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	2.43 ± 0.79	2.61 ± 0.85	3.07 ± 1.06	<0.001	2.64 ± 0.97	2.65 ± 0.87	2.77 ± 0.91	<0.001
Energy intake (Kcal/d)	1312.60 ± 312.38	1461.30 ± 352.71	1757.74 ± 426.67	<0.001	1509.69 ± 410.58	1463.66 ± 376.37	1516.36 ± 391.34	<0.001
Age group				<0.001				<0.001
35-44 years	47.1%	32.1%	29.4%		35.6%	36.1%	31.3%	
45-54 years	28.5%	31.9%	35.0%		27.4%	34.4%	33.4%	
55-64 years	24.4%	36.0%	35.6%		36.9%	29.5%	35.3%	
BMI group				<0.001				<0.001
underweight	1.6%	1.0%	0.9%		1.5%	0.9%	0.8%	
normal	52.5%	50.1%	49.6%		43.2%	52.6%	56.6%	
overweight	30.6%	33.3%	33.2%		33.1%	32.8%	31.5%	
obese	15.3%	15.6%	16.2%		22.2%	13.7%	11.0%	
Alcohol intake (0 or >20g/d)	10.6%	9.0%	9.5%	0.001	43.7%	0.0%	1.4%	<0.001
Fiber intake (<15g/1000Kcal)	99.0%	87.0%	44.1%	<0.001	76.7%	76.7%	84.9%	<0.001
Education level				<0.001				<0.001
≤9 years	29.3%	32.2%	27.7%		37.5%	25.8%	24.6%	
high school	38.9%	31.4%	28.9%		33.3%	33.9%	26.8%	
higher education	31.8%	36.5%	43.4%		29.1%	40.3%	48.6%	
Physical activity				<0.001				<0.001
inactive	22.0%	16.9%	12.0%		17.3%	16.3%	17.7%	
moderately inactive	32.4%	29.5%	23.9%		30.0%	27.9%	30.4%	
moderately active	27.1%	27.3%	26.8%		26.9%	27.4%	25.5%	
active	18.4%	26.3%	37.2%		25.8%	28.5%	26.4%	
Smoking status				<0.001				<0.001
non-smoker	48.6%	50.2%	53.9%		64.6%	48.1%	42.7%	
1-4 cigarette	2.6%	2.8%	2.4%		1.9%	2.6%	3.5%	
5-14 cigarette	12.7%	9.2%	5.2%		7.8%	9.3%	8.2%	
15-24 cigarette	4.7%	2.3%	0.9%		2.3%	2.2%	2.7%	
25+ cigarette	0.2%	0.1%	<1%		0.1%	<1%	0.1%	
occasional smoker	2.6%	2.9%	2.5%		0.9%	3.0%	4.0%	
former occasional smoker	8.3%	10.4%	12.1%		6.1%	11.6%	13.4%	
former smoker	20.3%	22.1%	22.9%		16.2%	23.1%	25.5%	
Medication for hypertension	11.4%	14.4%	14.9%	<0.001	16.4%	13.6%	13.1%	<0.001
Medication for dyslipidemia	2.0%	3.4%	3.4%	<0.001	3.9%	3.0%	2.8%	<0.001
IFG or IGT or Diabetes	18.7%	18.2%	17.6%	0.49	20.2%	17.6%	17.7%	<0.001
Diabetes	1.7%	1.8%	1.2%	0.008	2.1%	1.2%	1.3%	<0.001

***Table 6**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of dietary patterns in male participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables*

	"Mixed pattern" in men				"Refined food pattern" in men			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	2.40 ± 0.54	3.19 ± 0.67	4.81 ± 1.49	<0.001	2.93 ± 1.16	3.32 ± 1.10	3.99 ± 1.33	<0.001
Energy intake (Kcal/d)	1238.74 ± 256.44	1453.83 ± 312.33	1851.16 ± 432.14	<0.001	1726.21 ± 444.10	1933.81 ± 471.44	2402.09 ± 579.37	<0.001
Age group				<0.001				<0.001
35-44 years	14.4%	37.2%	52.7%		15.8%	35.9%	51.1%	
45-54 years	31.7%	32.1%	29.4%		26.7%	33.5%	30.7%	
55-64 years	53.9%	30.7%	17.9%		57.5%	30.5%	18.2%	
BMI group				<0.001				<0.001
underweight	1.2%	1.1%	1.2%		0.2%	0.2%	0.2%	
normal	52.1%	51.5%	47.8%		31.5%	34.3%	37.8%	
overweight	34.8%	32.4%	30.9%		51.2%	49.8%	47.1%	
obese	11.9%	14.9%	20.1%		17.0%	15.7%	14.9%	
Alcohol intake (0 or >20g/d)	12.1%	8.2%	9.4%	<0.001	8.5%	7.8%	10.3%	<0.001
Fiber intake (<15g/1000Kcal)	62.1%	81.4%	90.5%	<0.001	84.8%	97.4%	99.5%	<0.001
Education level				<0.001				<0.001
≤9 years	45.4%	28.8%	18.1%		45.8%	34.1%	30.7%	
high school	23.8%	33.9%	37.7%		28.4%	38.3%	43.0%	
higher education	30.8%	37.3%	44.2%		25.8%	27.6%	26.3%	
Physical activity				<0.001				<0.001
inactive	18.8%	16.8%	14.8%		18.0%	17.8%	17.7%	
moderately inactive	32.4%	29.4%	25.7%		27.7%	28.8%	28.7%	
moderately active	25.9%	27.6%	27.2%		27.6%	29.3%	30.7%	
active	22.8%	26.3%	32.3%		26.7%	24.0%	22.9%	
Smoking status				<0.001				<0.001
non-smoker	48.2%	50.5%	54.6%		45.7%	54.1%	57.6%	
1-4 cigarette	3.1%	2.7%	2.2%		2.8%	2.4%	2.8%	
5-14 cigarette	11.2%	8.9%	6.8%		5.8%	5.0%	5.3%	
15-24 cigarette	2.4%	2.6%	2.2%		3.2%	2.8%	3.5%	
25+ cigarette	0.1%	<1%	0.1%		0.4%	0.2%	0.4%	
occasional smoker	2.3%	2.6%	2.8%		2.4%	3.9%	3.5%	
former occasional smoker	10.2%	10.2%	10.9%		11.0%	9.9%	8.6%	
former smoker	22.6%	22.4%	20.4%		28.8%	21.6%	18.1%	
Medication for hypertension	16.9%	14.0%	11.4%	<0.001	22.6%	13.5%	8.6%	<0.001
Medication for dyslipidemia	3.7%	3.0%	2.5%	<0.001	10.3%	5.0%	2.9%	<0.001
IFG or IGT or Diabetes	19.5%	17.7%	17.3%	<0.001	23.6%	20.2%	18.1%	<0.001
Diabetes	1.7%	1.4%	1.2%	0.004	2.7%	1.6%	1.5%	<0.001

***Table 6**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of dietary patterns in male participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables (**continued**)*

	"Alcohol pattern" in men				"Prudent pattern" in men			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	3.27 ± 1.30	3.37 ± 1.15	3.60 ± 1.27	<0.001	4.14 ± 1.57	3.11 ± 0.98	3.27 ±1.07	<0.001
Energy intake (Kcal/d)	1986.75 ± 560.58	1959.25 ± 529.61	2082.27 ± 554.36	<0.001	2038.51 ± 576.13	1895.06 ± 497.10	2180.27 ± 578.88	<0.001
Age group				<0.001				<0.001
35-44 years	33.4%	40.2%	27.6%		49.1%	33.4%	24.2%	
45-54 years	30.5%	30.0%	31.6%		28.9%	31.4%	33.2%	
55-64 years	36.1%	29.7%	40.8%		21.9%	35.2%	42.6%	
BMI group				<0.001				<0.001
underweight	0.5%	0.2%	0.2%		0.3%	0.3%	0.2%	
normal	33.8%	34.4%	35.8%		29.7%	35.6%	36.7%	
overweight	46.3%	49.3%	51.4%		49.4%	50.0%	49.0%	
obese	19.4%	16.2%	12.6%		20.6%	14.1%	14.2%	
Alcohol intake (0 or >20g/d)	25.2%	0.0%	16.1%	<0.001	8.3%	7.3%	9.9%	<0.001
Fiber intake (<15g/1000Kcal)	93.0%	95.4%	96.5%	<0.001	99.8%	98.7%	81.0%	<0.001
Education level				<0.001				<0.001
≤9 years	44.3%	35.8%	28.5%		33.0%	38.5%	34.1%	
high school	35.0%	39.8%	32.5%		44.3%	35.5%	31.7%	
higher education	20.6%	24.3%	39.0%		22.7%	26.1%	34.3%	
Physical activity				<0.001				<0.001
inactive	17.8%	17.1%	19.5%		18.3%	18.4%	14.9%	
moderately inactive	27.7%	27.7%	30.8%		28.3%	29.3%	26.3%	
moderately active	29.7%	29.9%	26.5%		30.8%	29.4%	27.8%	
active	24.8%	25.3%	23.2%		22.6%	22.8%	31.0%	
Smoking status				<0.001				<0.001
non-smoker	60.6%	52.4%	45.7%		50.2%	53.5%	54.8%	
1-4 cigarette	2.0%	2.7%	3.2%		3.3%	2.5%	2.2%	
5-14 cigarette	5.7%	5.5%	4.9%		7.2%	4.8%	3.5%	
15-24 cigarette	3.1%	3.3%	3.3%		5.0%	3.1%	1.2%	
25+ cigarette	0.4%	0.2%	0.5%		0.6%	0.3%	0.2%	
occasional smoker	1.4%	3.7%	4.1%		4.3%	3.0%	2.5%	
former occasional smoker	7.3%	10.3%	11.3%		9.6%	9.3%	11.1%	
former smoker	19.5%	21.9%	27.0%		19.8%	23.2%	24.6%	
Medication for hypertension	16.5%	14.1%	14.6%	<0.001	11.0%	14.1%	17.0%	<0.001
Medication for dyslipidemia	6.7%	6.1%	6.1%	<0.001	4.0%	5.4%	7.9%	<0.001
IFG or IGT or Diabetes	22.3%	18.8%	20.3%	<0.001	19.3%	19.3%	20.7%	0.023
Diabetes	2.1%	1.7%	1.9%	0.15	1.9%	1.8%	2.0%	0.86

***Table 6**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of dietary patterns in male participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables (**continued**)*

	"White bread and sweet spreads pattern" in men			
	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	3.69 ± 1.36	3.22 ± 1.12	3.39 ± 1.19	<0.001
Energy intake (Kcal/d)	1959.82 ± 539.94	1883.92 ± 500.90	2266.42 ± 576.44	<0.001
Age group				<0.001
35-44 years	51.7%	31.8%	24.2%	
45-54 years	31.8%	30.7%	30.9%	
55-64 years	16.5%	37.5%	44.9%	
BMI group				<0.001
underweight	0.2%	0.1%	0.3%	
normal	32.3%	34.0%	39.4%	
overweight	49.6%	49.7%	47.0%	
obese	18.0%	16.2%	13.4%	
Alcohol intake (0 or >20g/d)	9.2%	7.5%	9.1%	<0.001
Fiber intake (<15g/1000Kcal)	97.3%	91.8%	97.4%	<0.001
Education level				<0.001
≤9 years	21.9%	35.6%	52.6%	
high school	44.9%	36.7%	29.6%	
higher education	33.2%	27.7%	17.8%	
Physical activity				<0.001
inactive	18.1%	16.8%	17.8%	
moderately inactive	26.4%	28.1%	30.1%	
moderately active	28.8%	28.7%	31.4%	
active	26.6%	26.4%	20.7%	
Smoking status				<0.001
non-smoker	56.8%	53.2%	48.9%	
1-4 cigarette	2.4%	2.5%	3.1%	
5-14 cigarette	4.2%	4.5%	7.1%	
15-24 cigarette	2.3%	2.7%	4.8%	
25+ cigarette	0.3%	0.3%	0.5%	
occasional smoker	4.7%	3.0%	2.5%	
former occasional smoker	10.4%	10.7%	8.2%	
former smoker	18.9%	23.1%	24.8%	
Medication for hypertension	11.4%	16.8%	14.2%	<0.001
Medication for dyslipidemia	4.5%	7.1%	5.3%	<0.001
IFG or IGT or Diabetes	17.4%	20.3%	22.8%	<0.001
Diabetes	1.5%	2.1%	2.4%	0.006

The highest quintiles of AP in both sexes, which have similar loadings, had lower proportions of obese individuals, never-smokers, individuals with lower alcohol and fiber intake, medication for hypertension, medication for dyslipidemia and dysglycemia cases but comprised higher proportions of more educated and older individuals. However, women in the highest quintile of AP are more physically active and have lower proportion of diabetes while men in the highest quintile of AP are less physically active. There was no difference in the proportion of diabetes case across quintiles of AP in men.

Finally, the highest quintiles of PP in men and PP in women comprised higher proportions of older, highly educated, physically active, never-smokers, individuals with higher fiber intake, and medications for hypertension and dyslipidemia. The highest quintile of PP in women was also characterized by higher proportion of individuals with obesity and lower proportions of individuals with high alcohol intake and diabetes while the corresponding quintile of PP in men comprised less obese individuals and more individuals who drink more and with dysglycemia case.

Associations between diabetes and dietary and carbon footprint patterns

In the present study the proportion of dysglycemia (IFG or IGT or Diabetes) among VIP participants was found to be 19.09% (men: 20.05% vs women: 18.19%) while the proportion of diabetes was 1.68% (men: 1.91% vs women: 1.46%).

Before performing univariate logistic regression on dietary component scores, we investigated the correlations (*r*) between dietary component scores. Except for the scores of dietary components 2 and 4 in men (RFP and PP, $r = -0.1822$), correlations between dietary component scores were weak for both men and women ($|r| < 0.15$) which justified the univariate logistic regression on component scores [supplementary tables **S2** and **S3**]. In crude analysis, dysglycemia in women was negatively associated with MP and AP while in male participants, it was negatively associated with MP, RFP and AP (AP: $P_{trend} = 0.126$), but positively associated with BSP. In another crude analysis, diabetes was negatively associated with MP, PP and AP in women; in men, it was negatively associated with RFP but positively associated with BSP.

Results from multivariate logistic regression showed that the third dietary component (PP) in women, which is characterized by higher intake of fruits, vegetables, high fiber cereals, vegetable oil and water, is significantly associated with dysglycemic state with $OR_{TC3\ Q5\ vs.\ Q1} = 0.82$ (95% CI 0.69—0.97, $P_{trend} = 0.003$). It was also significantly associated with diabetes with $OR_{TC3\ Q5\ vs.\ Q1} = 0.37$ (95% CI 0.17—0.78, $P_{trend} = 0.002$). The first dietary component (MP) in men, characterized by high intake of animal protein, fruits, vegetables, oil, pasta and high fiber cereals, potatoes and legumes, is significantly associated with dysglycemia ($OR_{TC1\ Q5\ vs.\ Q1} = 0.80$ (95% CI 0.65—0.97)) only for those in the 5th score quintile with no significant trend being observed across quintile levels ($P_{trend} = 0.075$) overall. When excluding prediabetic participants, the association between this component and diabetes was not significant, $OR_{TC1\ Q5\ vs.\ Q1} = 1.26$ (95% CI 0.63—2.51, $P_{trend} = 0.541$). MP in women was however significantly associated with diabetes for those in the 5th quintile with $OR_{TC1\ Q5\ v.\ Q1} = 0.35$ (95% CI 0.14—0.85). But this association did not show any significant trend across quintiles ($P_{trend} = 0.052$). However, after performing a separate logistic regression changing the reference quintile from the first to the fifth quintile, MP in women was significantly

associated with diabetes outcome: $OR_{TC1\ Q1\ vs.\ Q5}$=2.86 (95% CI 1.18—6.95); $OR_{TC1\ Q2\ vs.\ Q5}$=2.74 (95% CI 1.13—6.63); $OR_{TC1\ Q3\ vs.\ Q5}$=2.19 (95% CI 0.88—5.45); $OR_{TC1\ Q4\ vs.\ Q5}$=2.80 (95% CI 1.15—6.79) and P_{trend}= 0.052. No other significant association was observed between food patterns and the two disease outcomes. See [**Table 7** and **Table 8**] below.

***Table 7**. Odd Ratios of dysglycemic state (1=IFG/IGT/Diabetes, 0=Normoglycemic) and Quintiles of dietary patterns identified by Treelet Transform in women (n=36,770) and men (n=34,463) in the Västerbotten Intervention Program (VIP), 1996-2013*

	Q1	Q2		Q3		Q4		Q5		P_{trend}
	OR	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	
Women										
MP	Ref.	1.15	1.00—1.32	1.07	0.93—1.24	1.07	0.92—1.24	0.98	0.83—1.16	*0.570*
RFP	Ref.	0.98	0.85—1.14	0.87	0.75—1.01	0.91	0.78—1.06	0.91	0.77—1.08	*0.156*
PP	Ref.	0.92	0.80—1.06	0.90	0.78—1.03	**0.81**	0.69—0.94	**0.82**	0.69—0.97	***0.003***
AP	Ref.	1.01	0.86—1.19	1.06	0.91—1.24	1.03	0.88—1.20	1.08	0.93—1.26	*0.335*
Men										
MP	Ref.	0.91	0.77—1.07	0.93	0.79—1.10	0.95	0.80—1.13	**0.80**	0.65—0.97	*0.075*
RFP	Ref.	1.01	0.85—1.20	1.12	0.95—1.33	0.92	0.77—1.10	1.13	0.95—1.36	*0.468*
AP	Ref.	0.88	0.74—1.04	0.90	0.76—1.07	0.99	0.84—1.19	0.94	0.80—1.11	*0.951*
PP	Ref.	1.16	0.97—1.39	1.04	0.86—1.24	0.95	0.79—1.14	1.01	0.84—1.21	*0.242*
BSP	Ref.	0.93	0.78—1.12	0.89	0.74—1.07	0.92	0.77—1.10	0.88	0.73—1.04	*0.187*

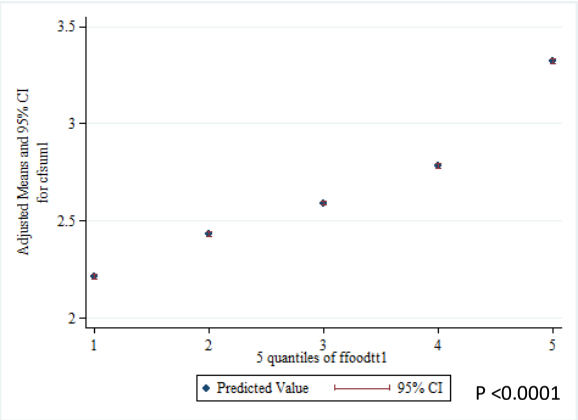
MP, mixed pattern; RFP, refined food pattern; PP, prudent pattern; AP, alcohol pattern; BSP, white bread sweet spreads pattern; Q1, quintile 1; Q2, quintile 2; Q3, quintile 3; Q4, quintile 4; Q5, quintile 5; Ref., reference category
*Models were adjusted for potential confounders: **total energy intake** in Kcal/day including from alcohol (continuous), **physical activity** (Cambridge index:1=Inactive[reference] 2=Moderately inactive 3=Moderately active 4=Active), **age group** (35-44 years[ref.], 45-54 years, 55-64 years), **education level** (≤9 years[ref.], 12 years, higher education), **smoking status** (Nonsmoker[ref.], 1-4 cigarette/d, 5-14 cigarette/d, 15-24 cigarette, ≥25 cigarette/d, occasional smoker, former occasional smoker, former smoker), **alcohol intake** in g/day(≤20[ref.], abstain/>20), **fiber intake** in g/1000Kcal (<15, ≥15[ref.]), **component scores** of other dietary patterns (continuous); potential effect modifiers: **body mass index** in Kg/m^2 (underweight: BMI <18.5; normal[ref.]: 18.5 ≤ BMI <25; overweight: 25 ≤ BMI <30; obese: BMI ≥30), **medication for hypertension** (yes, no[ref.]), **medication for dyslipidemia** (yes, no[ref.])*

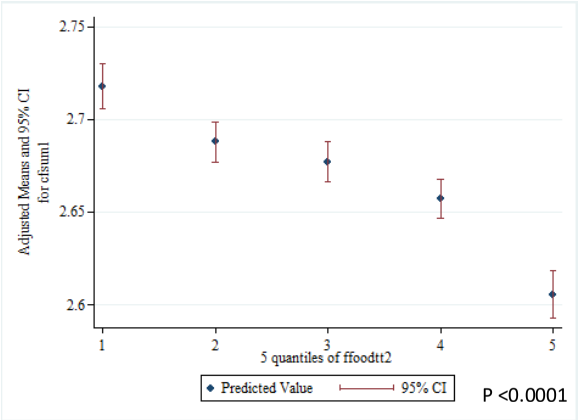
***Table 8.** Odd Ratios of Diabetes (1=Diabetes, 0=Normoglycemic) and Quintiles of dietary patterns identified by Treelet Transform for women (n=30,526) and men (n=28,089) in the Västerbotten Intervention Program (VIP), 1996-2013*

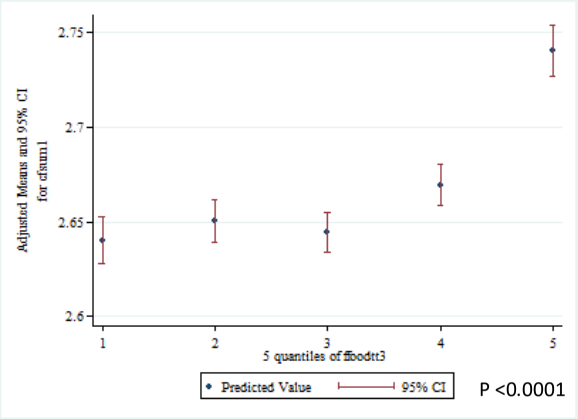
	Q1	Q2		Q3		Q4		Q5		P_{trend}
	OR	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	
Women										
MP	Ref.	0.96	0.55—1.67	0.76	0.41—1.43	0.98	0.53—1.80	**0.35**	0.14—0.85	*0.052*
RFP	Ref.	0.95	0.48—1.87	0.97	0.50—1.88	1.25	0.66—2.34	0.87	0.43—1.76	*0.982*
PP	Ref.	0.74	0.42—1.32	0.78	0.44—1.39	**0.40**	0.20—0.81	**0.37**	0.17–0.78	***0.002***
AP	Ref.	0.85	0.44—1.62	0.80	0.41—1.54	0.63	0.32—1.24	0.90	0.49—1.66	*0.587*
Men										
MP	Ref.	1.10	0.62—1.96	0.99	0.54—1.83	1.20	0.64—2.26	1.26	0.63—2.51	*0.541*
RFP	Ref.	1.03	0.57—1.84	0.65	0.32—1.29	1.29	0.71—2.32	1.15	0.60—2.19	*0.521*
AP	Ref.	0.90	0.47—1.72	0.73	0.37—1.45	1.45	0.80—2.63	1.04	0.56—1.93	*0.404*
PP	Ref.	0.94	0.51—1.76	0.64	0.33—1.26	0.75	0.39—1.44	0.90	0.47—1.70	*0.579*
BSP	Ref.	0.66	0.30—1.44	1.09	0.55—2.13	0.98	0.50—1.94	1.31	0.68—2.51	*0.189*

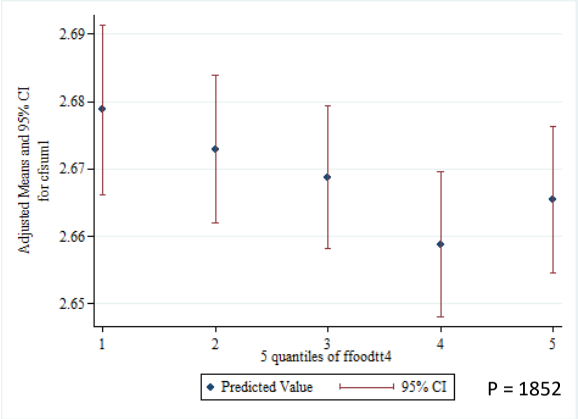
MP, mixed pattern; RFP, refined food pattern; PP, prudent pattern; AP, alcohol pattern; BSP, white bread sweet spreads pattern; Q1, quintile 1; Q2, quintile 2; Q3, quintile 3; Q4, quintile 4; Q5, quintile 5; Ref., reference category
*Models were adjusted for potential confounders: **total energy intake** in Kcal/day including from alcohol (continuous), **physical activity** (Cambridge index:1=Inactive[reference] 2=Moderately inactive 3=Moderately active 4=Active), **age group** (35-44 years[ref.], 45-54 years, 55-64 years), **education level** (≤9 years[ref.], 12 years, higher education), **smoking status** (Nonsmoker[ref.], 1-4 cigarette/d, 5-14 cigarette/d, 15-24 cigarette, ≥25 cigarette/d, occasional smoker, former occasional smoker, former smoker), **alcohol intake** in g/day(≤20[ref.], abstain/>20), **fiber intake** in g/1000Kcal (<15, ≥15[ref.]), **component scores** of other dietary patterns (continuous); potential effect modifiers: **body mass index** in Kg/m^2 (underweight: BMI <18.5; normal[ref.]: 18.5 ≤ BMI <25; overweight: 25 ≤ BMI <30; obese: BMI ≥30), **medication for hypertension** (yes, no[ref.]), **medication for dyslipidemia** (yes, no[ref.])*

We also computed the adjusted mean dietary GHGe [in Kg CO_2e/day] by quintiles of dietary components. Accordingly, for women, dietary GHGe showed statistically significant increasing trends in MP and PP but a statistically significant decreasing trend in RFP. For men, the observed statistically significant trends of GHGe across quintiles of food patterns were increasing in MP and AP, and decreasing in PP and BSP [**Figure 4** and **Figure 5**]. In crude analysis dietary GHGe showed statistically significant increase in trend across quintiles of RFP for both men and women and across quintiles of AP for women while the trends in the other dietary patterns were in the same direction as in the case of the adjusted analysis [see tables **5 & 6**].





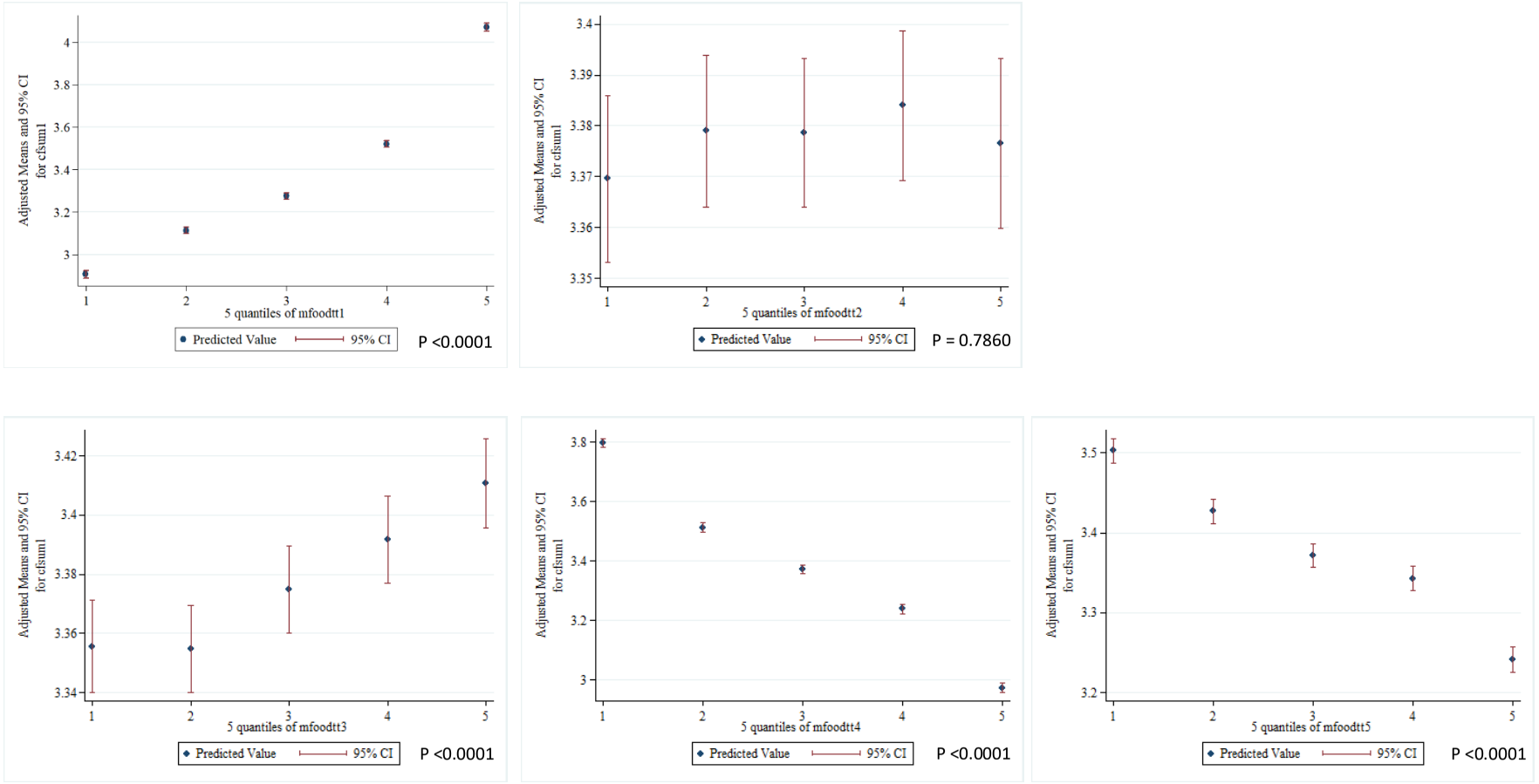




Legend:

ffoodtt1: variable name for mixed pattern
ffoodtt2: variable name for refined food pattern
ffoodtt3: variable name for prudent pattern
ffoodtt4: variable name for alcohol pattern

Figure 4.. *Adjusted mean values and 95% confidence interval of GHGe in Kg CO_2e/day across quintiles of dietary patterns in female participants of Västerbotten Intervention Programme (VIP), 1996-2013*



Legend:

mfoodtt1: variable name for mixed pattern
mfoodtt2: variable name for refined food pattern
mfoodtt3: variable name for alcohol pattern
mfoodtt4: variable name for prudent pattern
mfoodtt5: variable name for white bread and spreads pattern

Figure 5. *Adjusted mean values and 95% confidence interval of GHGe in Kg CO_2e/day across quintiles of dietary patterns in male participants of Västerbotten Intervention Programme (VIP), 1996-2013*

Finally, crude and adjusted analyses of proportions of the two outcomes (dysglycemia and diabetes) across quintiles of GHGe component scores were performed separately. Results from the analyses are shown on supplementary tables **S4** and **S5**. Quintiles of carbon footprint (GHGe) were not significantly associated with either of the outcomes after adjusting for confounders except for *cfFDB* in men which was significantly associated with proportions of dysglycemia (Q_1 = 18.9% *v* Q_5 = 17.2%; P = 0.0351). Crude analyses showed that prior to adjusting for confounders, the "dysglycemia" outcome had statistically significant different proportions across quintiles of GHGe component scores for both men and women except for *cfPP* in men (P = 0.5600); meanwhile, the outcome "Diabetes" had statistically significant positive association with *cfPP* in women (P = 0.0346) and statistically significant negative association with *cfRFP* in men (P < 0.0001), *cf*MP (P = 0.0111), *cf* RFP (P = 0.0118) and *cf* AP (P = 0.0001) in women.

Discussion

The present study is based on VIP register, one of the most comprehensive registers with over 74,000 unique records between 1996 and 2013, giving us strong statistical power while conducting pattern analysis on the dietary and *cf* data. We have employed TT to identify dietary and carbon footprint patterns which are correlated and mostly have structural similarities. Relating these identified patterns to dysglycemia and diabetes prevalence showed that prudent dietary pattern in women was inversely associated with these disease outcomes in dose-response manner. An evaluation of emissions from these dietary patterns, however, showed that higher adherence to prudent dietary pattern in women resulted in higher dietary emission. A further analysis investigated if the emission patterns, which are correlated with dietary patterns, showed evidence for association with dysglycemia and diabetes. None of the identified emission patterns were associated with dysglycemia and diabetes indicating that their usage as a proxy for dietary patterns may not be appropriate.

Regarding dietary patterns:

The significance of dietary data dimension-reduction for evaluation of diet—disease association has been shown in several studies [6,23,24,59,60]. TT is one of such dimension-reduction methods which was developed to improve interpretation of patterns through sparsity. Since the publication of the methodology by Lee et al. [17], TT has been used only a handful times in the nutrition and health sphere [19-21] for data reduction, and only one of them concerned dietary patterns and T2DM [21]. By using TT, we had obtained four dietary patterns for women and five dietary patterns for men. Most researches report two main patterns: one mostly labelled as "unhealthy or western" characterized by high loading for processed meat, red meat, refined foods, sweets and sugary beverages; another mostly labelled as "healthy or prudent" characterized by high loading for vegetables, fruits, high fiber grains, legumes, low fat dairy, fish and poultry. Despite this popular labeling method, there are variations in the data reduction methods applied, and in the food-items included in the labelled patterns. In our study, patterns which resemble these labels have been retained. However, neither high fat nor low fat dairy loaded on the dietary components. In our study,

the food items included in these groups were diverse (high fat dairy including sour milk, yoghurt and milk with 3% fat content and low-fat dairy including sour milk, yoghurt and milk with fat content of 0.5-1.5%). The variation in intake might, hence, be small to be detected at the optimal cut levels. Moreover, a comparison of our study with another study conducted on the VIP data [46] showed that the "refined food pattern" in women resembled the "coffee and sandwich" pattern described in that study. However, food groups such as ice-cream, soda and sweets had loaded on the "refined food pattern" additionally while coffee was absent; Furthermore, there were several food items from the "High Fat" pattern in women which were recovered in the "refined food pattern" in our study. The "Fruit and Vegetable" pattern in women from that study was similar to the combination of "mixed" and "prudent" patterns in our study. In men, the "Fruit and Vegetable" from that study resembled our "mixed pattern"; Food items listed in the "Tea, Soda and Cookies" and "High Fat "were included in "refined food pattern" in our study. Besides that, there were major differences in the dietary patterns described in the two studies. In our study, more dietary patterns were defined for men than women while in the study by Winkvist et al. [46] more patterns were identified among women. Another difference between the two studies is that our study was able to identify a pattern solely from alcohol intake. This resulted from the multi-scale nature of TT which creates sparse component loadings and facilitates detection of localized sources of variations. The difference in dimension reduction methodology used (cluster analyses vs TT) can partly explain the difference in the number and structure of the identified patterns, as it has been shown in previous studies which compared results from various methods [61,62]. Another plausible explanation is the time difference between the study population. While our study included unique health examinations between 1996-2013, Winkvist et al. [46] had used data from 1992-2005. This means while we used food intake measurements from the short FFQ entirely, the dietary data in the study by Winkvist et al. [46] was harmonized when combining data from the long FFQ (before 1996) with data from the short FFQ. This time difference also may result population wide dietary habit change as noted in another study by the same research group based on the VIP [63]. Other external factors, such as the 2005 Swedish dietary guideline [64] may also have differential effect on dietary preference. Consequently, by performing both methods on the same population, it will be possible to verify if the identified patterns are close to the intrinsic true dietary patterns in the VIP register [65].

The total variance explained by the dietary components was 21.42% for women and 24.96% for men. This is in line with other studies which have used food groups for derivation of dietary patterns [18, 20; 66]. It is known that the total number of food groups introduced to the data reduction method inversely affects the total percentage of variance explained by each component [67].

Concerning the association between the identified dietary patterns and disease outcomes (dysglycemia or Diabetes), the "prudent pattern" in women had significant inverse association with both disease outcomes, its association with diabetes being more negative. "prudent pattern" in men was not significantly associated with either of the outcomes. It is important to note that despite having the same labels, the two patterns had important differences. First, "prudent pattern" in women had stronger positive correlation with vegetable, fruit, fresh fish, fiber, whole grain and water intakes, but stronger negative correlation with sweets and sugary drinks when compared to "prudent pattern" in men. It has

been previously shown that diet rich with vegetable, fruit, fiber and unsaturated fats may be protective against T2DM [68,69]. Second, the structure between the two "prudent patterns" differ in loaded food groups. "prudent pattern" in men had negative loading from pasta, legumes, processed and red meat, poultry and fish. Structurally, it resembled more of a vegetarian pattern. Also, water had not loaded on this pattern. The significance of water in prevention of T2DM had been reported previously both considering water intake per se and substitution of sugary drinks with water [70-72]. Third, it is possible that male participants might have reported their intention to eat more fruits, vegetables and high fiber cereals instead of their actual food intake. The significant negative association of "prudent pattern" in women with T2DM in our study was not found in the only study which used TT to identify dietary patterns [21]. This contradiction probably arises from the food grouping technique and the structural difference of the similarly labelled patterns a reflection of the variation in dietary preferences which exists among different population sets. The authors [21] concluded that "*factor analysis is more appropriate to identify overall dietary patterns associated with diabetes when compared with treelet transform analysis*". Our study does not support that claim. The inconsistency regarding dietary pattern analysis and disease outcome is well known [22,23]. While other researches which have not used TT, such as the study by Liu et al. [73] have reported western type of dietary pattern rich in refined food may increase the odds of having diabetes similar to the factor analysis in the study by Schoenaker et al. [21], our analysis on "refined food pattern" does not show such result. In the contrary, in [21], neither factor analysis nor TT shows significant association of prudent pattern with diabetes while we have reported prudent pattern is inversely associated with diabetes. However, these heterogenous results do not mean that one dimension reduction technique is superior over the other. The lack of association with a disease outcome does not render a pattern analysis invalid as it can still capture truly existing dietary patterns among a given population [74]. A concurrent pattern analysis with another dimension reduction method can be used to validate the identified patterns [65].

Regarding cf patterns

It is only recently that FFQs started to be used for dietary *cf* calculation, a shift from previous researches which focused on *cf* of individual food items which are known to have big contribution to dietary *cf* [28,40,42]. In our study, total daily mean value of *cf* for the VIP participants was 3.02 $KgCO_2e$. This finding is in line with other studies in the Swedish setting [40,75,76]. The study by Sjörs et al. [40] had reported a mean value of 3.76 Kg CO_2e/day. However, that study had much smaller sample size comprising younger participants. Also, the FFQ which is used in our study was shorter than the FFQ used in the study by Sjörs et al. [40].

By using the same dimension reduction method (TT) as in the case of dietary pattern analysis, we have identified four *cf* patterns in women and five *cf* patterns in men explaining 21.52% and 25.37% of the total variance respectively. Our motivation for this *a posteriori* approach is that previous estimates of daily dietary emission do not take variations of *cf* patterns across a population into consideration. By performing pattern analysis, we could identify truly existing *cf* patterns in the population (VIP participants).

Most of the identified *cf* patterns strongly reflected the dietary patterns with higher degree of correlation and similar loadings of food items except the fifth dietary ("white bread sweet spreads", BSP) and *cf* (*cf* from "fatty dairy and butter", *cf*FDB) patterns in men. A sensitivity

analysis had shown *cf*FDB was not stable and hence its difference from BSP. Moreover, the calculated values of mean *cf* estimate across quintiles of *cf* patterns were comparable with mean *cf* estimates across quintiles of dietary patterns indicating that there was agreement between quintiles of the dietary and *cf* patterns.

Despite these similarities, it is important to note some observed cross-loadings of fast-food, salty snacks and their corresponding emission on "mixed" dietary pattern and "*cf* from refined food pattern, *cf*RFP*"* in women. The moderate correlation between "*cf* from prudent pattern", *cf*PP with "prudent" dietary pattern in women was due to absence of emission values of vegetable, water and oil from (and addition of emissions from potatoes and low-fat dairy to) *cf*PP and this may have rendered its association with diabetes (or prediabetes) insignificant. Moreover, the small reduction of IFG/IGT/Diabetes prevalence noticed across *cf*FDB may be due to residual confounding such as interactions with BMI. It is also noted above that this pattern was not stable rendering its use in regression model not appropriate. Adjusted mean *cf* estimates in the highest quintile of prudent pattern in women was larger compared to *cf* values in the lowest quintile. It was also larger when compared to the *cf* estimates in the highest quintiles of "alcohol pattern" and "refined food pattern" indicating while the prudent pattern may be protective from diabetes, high adherence to it may not have mitigating effect when it comes to emission.

Moreover, we have not identified in our study any *cf* pattern which is associated with diabetes when adjusting for other confounders. Other researches have also found similar results showing that healthy dietary patterns do not necessarily have lower emission or patterns with higher emission are not necessarily unhealthy, depending on the measured health outcome [34]. In the contrary, others have found moderate positive association between healthy diets and lower emission [36,77]. Considering this, despite similarities between dietary and *cf* patterns, using *cf* patterns for disease prediction may not be suitable, and its application should be further investigated.

In the current study we have used carbon footprints as measurement of sustainability as LCA data on dietary emission are readily available. However, there are also other environmental measurements for sustainability which are not considered here, such as land use, yield, water footprint and biodiversity which have both competing and positive synergies amongst them.

Methodological Consideration and Limitations

In the present study, first, we have linked the DietVIP short FFQ with dietary *cf* expressed as Kg Carbon dioxide equivalents (Kg CO_2e) per Kg of the food items available in the FFQ based on LCA data reported by Elin Röös, Swedish University of Agricultural Science [49]. The *cf* of each food group was, then, obtained by summing *cf* of food items constituting that food group.

Second, we have identified dietary and *cf* patterns from the daily dietary intake and *cf* values respectively for both women and men. In fact, to the best of our knowledge, it is the first study to identify *cf* patterns *a posteriori* directly from FFQ, to compare them with the extracted dietary patterns and to use them for disease outcome evaluation, shifting the focus from known *cf* contributors to data-driven population-based *cf* patterns.

The cross-sectional design of the study means causality and temporality cannot be stated surely. In this spirit, it is advised to read the results with care. The results from this study can be used for hypothesis setting in other stronger researches. We have also followed thorough methodology to avoid the known pitfalls coming with such observational studies. Before the start of analysis, we had created some exclusion criteria which attenuate the problem related to reverse causality. For example, we have excluded individuals with known diabetes case. Our analysis has included individuals who have reported medication for hypertension and dyslipidemia, a scenario affecting dietary preference. As a result, we have adjusted for several confounders associated with diabetes which are available in the literature. We had also performed univariate regression and stepwise addition to explore whether the included variables were associated with the disease outcome. Inclusion of big number of covariates in regression model might result sparse data bias. We have developed our regression model by assessing events per variable (EPV). Because the covariates were added for confounder adjustment, we have rendered five EPV as sufficient as it is suggested by Vittinghoff et al. [78]. When we had less than 5 EPV, like in the final regression models of dietary patterns for women smokers of 25+ cigarettes/day, we had done separate analysis by combining these categories with adjacent categories, this however did not change the odd ratios or their 95% confidence intervals. However, because we have not included important confounders (such as family history of diabetes or hormone therapy status) in the analysis, we cannot rule out residual confounding.

Underreporting is also another problematic aspect when collecting data especially with self-administered FFQ. Previous research has shown that there is high under-reporting of calorie intake within the VIP [79]. This creates reporting bias both for the dietary data and *cf* data, as the *cf* values are based on dietary intake. The issue of under-reporting is also a well-known problem in studies focusing on dietary emission [28].

The long timespan in this study has given us large study population set. The use of almost the same FFQ for data collection between 1996 and 2013 possibly does not capture dietary shifts and introduction of new food items into the culinary. Having this large timespan, we are also assuming that emission estimates of food items have remained constant. Naturally, the next step will be conducting a validation study for the FFQ which investigates its validity for dietary emission studies, a big potential with indication for public health and environmental policy.

Another important aspect to consider is we have identified comparable dietary and *cf* patterns. One of the strength of our study is that we have mostly used the same source for the *cf* estimations. Elin Röös [49] has done thorough review of existing literature focusing on the Swedish setting and have compiled data based on similar LCA or system boundaries. The *cf* values given in the report include emissions resulting from primary production up to the packaging stage (the final form the food item is sold in the market). Emissions from transport have been taken into account explicitly for fruit and vegetables. Emphasis has been given on whether these food items are imported from Nordic countries, southern European Union (EU) or further away. For meat products, transport to Sweden was not considered because the transport emission covers very small percentage of the total emission. For other food items, Sweden was considered highly self-efficient. Due to these boundaries, which are set at the purchasing point we have excluded emission estimates from cooking and waste. Also, the FFQ measures cooked foods which means some unavoidable weight changes from cooking

processes are not considered when calculating emission estimate. So, it is important to read the result from this study with caution regarding dietary *cf* estimates. It is well known that it is difficult to calculate emissions from food purchase, preparation and waste. Several researches include rough estimates of these stages uniformly to different food items after calculating the total amount of emission while others use *cf* estimation from cradle to retail phase [28,36,40,41,80]. Formulating questions which address waste, energy use, distance travelled for grocery shopping, means of transport when shopping and purchasing preference (such as imported vs. locally produced vs. organic) can improve the existing FFQ's ability to capture dietary emissions. But it is important to note that the *cf* patterns retained in our study were mostly strongly correlated to the corresponding dietary patterns.

The application of TT for dietary data reduction is relatively new when compared to other well applied empirical methods, such as PCA and Cluster analysis. TT's strength is its quantitative technique of pattern extraction and its interpretability. The concerns which are raised regarding TT specifically are whether 1) sparsity of the identified patterns affects disease prediction and 2) sparsity achieved by TT is useful when adjusting for confounders by using patterns. Gorst-Rasmussen et al [81] argue that the ultimate goal of any dimension reduction method is creating latent variables which express variations. TT avoids the subjectivity in PCA which comes at the stage when a researcher uses rotation and loading truncation to create sparsity. Moreover, in addition to producing comparable patterns with PCA or Cluster analysis, Researches have shown the applicability of these patterns in Epidemiology studies. Out of the previous three publications on TT patterns, two have reported association of TT patterns with disease outcomes [18-20]. Our study promotes the usefulness of TT in nutrition epidemiology. If we want to adjust for confounding by pattern like the study by Imamura et al. [82], sparsity may not be useful because we want those patterns to be loaded by all variables. In our study, by investigating the correlation of the 34 food groups with AP in men and women we had found that oil, vegetable and salty snacks have stronger (but still weak) association with AP compared to other food groups. However, since specific dietary consumption is not used for stratification (such as drinkers vs. abstainers), we have not investigated the effect of confounding by pattern. We have stated before that TT components are not entirely uncorrelated. For this reason, when using quintiles of a component as main predictor, we opted to include component scores of other patterns (continuous variables) in the final regression model.

It is important, though, to note that like any of diet pattern analysis techniques, whether *a priori or a posteriori*, TT has its limitation. First, the researcher decides how much she trades off between interpretability and sparsity, meaning the cut-level. Cutting the tree near the leaves results sparse components creating zero loading for many of the food items; cutting the tree near the root results components similar to PCA, rendering interpretation difficult. This issue is solved by cross-validation technique which computes the optimal cut level for a given number of components. But, this results the second issue: how many components to retain. In our study, we first set a cut level at the root, and investigated the variances and the scree plot. Afterwards, we retained those components with variance >1.0, and included them in the cross-validation.

Finally, in our study, we have used quintiles (i.e. 5 quantiles) of component scores for evaluation of pattern—diabetes association. The use of quantiles in regression model is

debatable [83]. The main criticism which is raised by the opponents of this approach is that the dose-response effect is assumed constant for all continuous values that quantile was based on. Another point of criticism is that quantile cut-off values depend on the data i.e., it is not possible to compare results from different researches. Finally, as the comparison is done between two quantiles, the data from the other quantiles remains irrelevant. In the present study we have created quintiles based on the component scores for the regression analysis. While we acknowledge the points raised against quantiles, we also believe using quantiles in pattern analysis has its own advantages. First the use of pattern analysis in our study is to identify dietary and *cf* patterns in the population. These patterns are labelled based on their composition. The prudent pattern is associated with good health outcome and the western or refined food pattern is associated with unfavorable health outcome. Beyond that, pattern scores do not have exact threshold values where a significant disease odd occurs. One of the important element of epidemiology is conveying the result in understandable way, and the usage of quantile makes it easier. By comparing quantiles (that is degree of adherence), we can report odds and risks. Additionally, our study's aim is to simply estimate approximation of the rate ratio of diabetes and the use of quantiles suffices for this purpose. Second, the test for trend is important because it takes all quantiles into consideration. Moreover, performing separate analysis by changing the reference quantile to the other extreme helps the researcher to investigate the model further. In our case we have shown that by changing the reference quintile to the highest quintile of the MP in women, we had shown the 1st, 2nd and the 4th quintiles had higher ORs versus the 5th quintile. But the test of trend does not depend on which category is the reference, hence it remains constant. The significance level from this test can be used to investigate if there is a true trend across the quantiles.

Conclusion

The current study linked the dietary data from VIP register to LCA data of food items and reported results from dietary and *cf* pattern analyses in VIP participants by using a relatively new dimension reduction method, TT. Furthermore, it investigated the association of these patterns with dysglycemia and diabetes. Through this detailed approach, it is shown that empiric dietary patterns and diet-related emission patterns have moderate to strong structural similarities leading to similar arbitrary labelling. However, our study also showed that despite this similarity, some of the missing loading variables in emission patterns might render their usage as a proxy for dietary patterns inappropriate for diabetes risk evaluation. In our study, higher adherence to the prudent dietary pattern in women reduced the odds of having both dysglycemia and diabetes while *cf* patterns were not associated with either of the disease outcomes. By performing concurrent pattern analyses on both diet and diet related emissions, we have found that a healthy diet may not necessarily reduce GHGE. Finally, we have noted some of the methodological pitfalls. Future works might focus on validation of the DietVIP FFQ for dietary emission, designing of important emission related measurement variables within the VIP questionnaire, performing longitudinal studies to assess the predictive ability of dietary and emission patterns on the development of dysglycemia, and conducting dietary intervention trials which also measure changes in diet-related emission. Such future works provide high grade evidence for adoption of new methods by public health interventions, such as VIP. One could suggest that due to climate change and the increasing burden of life style-related noncommunicable diseases, it is important to develop methods that promote public health within healthy planetary systems.

Ethical consideration

The VIP data used for the study was obtained in encrypted form from the Department of Biobank, Umeå University after joint submission of application by the author of this thesis and the supervisor and after signing a data transfer agreement. Additionally, the study was approved by the regional ethics board [Dnr 08-131M, amendment Dnr 2017-235-32] for inclusion of researches on ecological footprints in the VIP. Further ethical approval was not necessary as the student thesis uses already ethically collected data.

Competing Interests

The author declares no competing interests.

Acknowledgment

The author thanks his supervisors Joacim Rocklöv and Margareta Norberg for their tremendous patience, guidance and encouragement. Ingegerd Johansson, from the department of odontology, was also helpful whenever the author needed additional data.

Special thanks to the Swedish Institute (**SI**) for the generous scholarship for the master's program in Public Health, Umeå University, and to the staff at the department of public health science for their assistance.

References

1. International Diabetes Federation. *IDF Diabetes Atlas, 8th edn.* Brussels, Belgium: International Diabetes Federation, 2017. http://www.diabetesatlas.org
2. Feldman, Adina L., et al. "Change in lifestyle behaviors and diabetes risk: evidence from a population-based cohort study with 10 year follow-up." *International Journal of Behavioral Nutrition and Physical Activity* 14.1 (2017): 39.
3. Ley, Sylvia H., et al. "Prevention and management of type 2 diabetes: dietary components and nutritional strategies." *The Lancet* 383.9933 (2014): 1999-2007.
4. Forouzanfar, Mohammad H., et al. "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013." *The Lancet* 386.10010 (2015): 2287-2323.
5. Diabetes Prevention Program Research Group. "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin." *New England journal of medicine* 346.6 (2002): 393-403.
6. Jannasch, Franziska, Janine Kröger, and Matthias B. Schulze. "Dietary Patterns and Type 2 Diabetes: A Systematic Literature Review and Meta-Analysis of Prospective Studies–3." *The Journal of nutrition* 147.6 (2017): 1174-1182.
7. Pan, An, et al. "Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis–." *The American journal of clinical nutrition* 94.4 (2011): 1088-1096.
8. Merlotti, C., A. Morabito, and A. E. Pontiroli. "Prevention of type 2 diabetes; a systematic review and meta-analysis of different intervention strategies." *Diabetes, Obesity and Metabolism* 16.8 (2014): 719-727.

9. Salmeron, Jorge, et al. "Dietary fiber, glycemic load, and risk of non—insulin-dependent diabetes mellitus in women." *Jama* 277.6 (1997): 472-477.
10. Liu, Simin, et al. "A prospective study of whole-grain intake and risk of type 2 diabetes mellitus in US women." *American journal of public health* 90.9 (2000): 1409.
11. Salmerón, Jorge, et al. "Dietary fiber, glycemic load, and risk of NIDDM in men." *Diabetes care* 20.4 (1997): 545-550.
12. Hu, Frank B., R. M. Van Dam, and S. Liu. "Diet and risk of type II diabetes: the role of types of fat and carbohydrate." *Diabetologia* 44.7 (2001): 805-817.
13. Guthrie, H. A., and J. C. Scheer. "Validity of a dietary score for assessing nutrient adequacy." *Journal of the American Dietetic Association* 78.3 (1981): 240-245.
14. Kennedy, Eileen T, et al. "The healthy eating index: design and applications." *Journal of the Academy of Nutrition and Dietetics* 95.10 (1995): 1103-1108.
15. Sacks, Frank M., et al. "Rationale and design of the Dietary Approaches to Stop Hypertension trial (DASH): a multicenter controlled-feeding study of dietary patterns to lower blood pressure." *Annals of epidemiology* 5.2 (1995): 108-118.
16. Trichopoulou, Antonia, et al. "Diet and overall survival in elderly people." *Bmj* 311.7018 (1995): 1457-1460.
17. Trichopoulou, Antonia, et al. "Adherence to a Mediterranean diet and survival in a Greek population." *New England Journal of Medicine* 348.26 (2003): 2599-2608.
18. Lee, Ann B.; Nadler, Boaz; Wasserman, Larry. "Treelets—an adaptive multi-scale basis for sparse unordered data." *The Annals of Applied Statistics* 2.2 (2008): 435-471
19. Gorst-Rasmussen, Anders, et al. "Exploring dietary patterns by using the treelet transform." *American journal of epidemiology* 173.10 (2011): 1097-1104.
20. Assi, Nada, et al. "A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor-defined breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC)." *Public health nutrition* 19.2 (2016): 242-254.
21. Schoenaker, Danielle AJM, et al. "Factor Analysis Is More Appropriate to Identify Overall Dietary Patterns Associated with Diabetes When Compared with Treelet Transform Analysis–3." *The Journal of nutrition* 143.3 (2013): 392-398.
22. Slattery, Martha L. "Analysis of dietary patterns in epidemiological research." *Applied physiology, nutrition, and metabolism* 35.2 (2010): 207-210.
23. McEvoy, Claire T., et al. "A posteriori dietary patterns are related to risk of type 2 diabetes: findings from a systematic review and meta-analysis." *Journal of the Academy of Nutrition and Dietetics*114.11 (2014): 1759-1775.
24. Maghsoudi, Zahra, Reza Ghiasvand, and Amin Salehi-Abargouei. "Empirically derived dietary patterns and incident type 2 diabetes mellitus: a systematic review and meta-analysis on prospective observational studies." *Public health nutrition*19.2 (2016): 230-241.
25. Lagerberg Fogelberg, Charlotte. *Towards environmentally sound dietary guidelines*. 2013.
26. Behrens, Paul, et al. "Evaluating the environmental impacts of dietary recommendations." *Proceedings of the National Academy of Sciences* (2017): 201711889.
27. Myhre, Gunnar, et al. "Anthropogenic and natural radiative forcing. " *Climate change* 423 (2013): 658-740.
28. Hoolohan, Claire, et al. "Mitigating the greenhouse gas emissions embodied in food through realistic consumer choices." *Energy Policy* 63 (2013): 1065-1074.
29. Vermeulen, Sonja J., Bruce M. Campbell, and John SI Ingram. "Climate change and food systems." *Annual Review of Environment and Resources* 37 (2012): 195-222.
30. Carlsson-Kanyama, Annika. "Climate change and dietary choices—how can emissions of greenhouse gases from food consumption be reduced?" *Food policy* 23.3-4 (1998): 277-293.
31. Flysjö, Anna Maria. *Greenhouse gas emissions in milk and dairy product chains: improving the carbon footprint of dairy products*. Aarhus Universitet, Institut for Agroøkologi, 2012.
32. Cederberg, Christel. *Greenhouse gas emissions from Swedish production of meat, milk and eggs 1990 and 2005*. SIK Institutet för livsmedel och bioteknik, 2009.
33. Sonesson, Ulf, Jennifer Davis, and Friederike Ziegler. *Food production and emissions of greenhouse gases: an overview of the climate impact of different product groups*. SIK Institutet för livsmedel och bioteknik, 2010.

34. Payne, Charlotte LR, Peter Scarborough, and Linda Cobiac. "Do low-carbon-emission diets lead to higher nutritional quality and positive health outcomes? A systematic review of the literature." *Public health nutrition* 19.14 (2016): 2654-2661.
35. Aleksandrowicz, Lukasz, et al. "The impacts of dietary change on greenhouse gas emissions, land use, water use, and health: a systematic review." *PloS one* 11.11 (2016): e0165797.
36. Milner, James, et al. "Health effects of adopting low greenhouse gas emission diets in the UK." *BMJ open* 5.4 (2015): e007364.
37. Huseinovic, E., et al. "Does diet intervention in line with nutrition recommendations affect dietary carbon footprint? Results from a weight loss trial among lactating women." *European journal of clinical nutrition* 71.10 (2017): 1241.
38. Sjörs, Camilla, et al. "Adherence to dietary recommendations for Swedish adults across categories of greenhouse gas emissions from food." *Public health nutrition* 20.18 (2017): 3381-3393.
39. Bälter, Katarina, et al. "Is a diet low in greenhouse gas emissions a nutritious diet? – Analyses of self-selected diets in the LifeGene study." *Archives of Public Health* 75.1 (2017): 17.
40. Sjörs, Camilla, et al. "Diet-related greenhouse gas emissions assessed by a food frequency questionnaire and validated using 7-day weighed food records." *Environmental Health* 15.1 (2016): 15.
41. Biesbroek, Sander, et al. "Does a better adherence to dietary guidelines reduce mortality risk and environmental impact in the Dutch sub-cohort of the European Prospective Investigation into Cancer and Nutrition?." *British Journal of Nutrition* 118.1 (2017): 69-80.
42. Veeramani, Anastasia, Goretty M. Dias, and Sharon I. Kirkpatrick. "Carbon footprint of dietary patterns in Ontario, Canada: A case study based on actual food consumption." *Journal of cleaner production* 162 (2017): 1398-1406.
43. Norberg, Margareta, et al. "The Västerbotten Intervention Programme: background, design and implications." *Global health action* 3.1 (2010): 4643.
44. Lonnberg, Goran. Västerbotten Intervention Programme. ”VHU 1990-2016” [INTERNET] (2017). https://public.tableau.com/profile/epiglobe#!/vizhome/VHU1990-2014/Adderatdeltagande. Accessed: 2018-05-02. Archived at http://www.webcitation.org/6z7hMpOWs
45. Johansson, Ingegerd, et al. "Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort." *Public health nutrition* 5.3 (2002): 487-496.
46. Winkvist, Anna, et al. "More distinct food intake patterns among women than men in northern Sweden: a population-based survey." *Nutrition journal* 8.1 (2009): 12.
47. Department of Biobank research. “NSDD- Basic Information” [INTERNET]. Umeå University. http://www.biobank.umu.se/biobank/northern-sweden-diet-database/basic-information/ Accessed: 2018-05-02. Archived at http://www.webcitation.org/6z7jm7GtD
48. Livsmedelsdatabasen (The National Food Agency food database, version 2017-12-15) [Internet]. Livsmedelsverket (The Swedish National Food Agency), Uppsala. 2017. http://www.livsmedelsverket.se/en/food-and-content/naringsamnen/ livsmedelsdatabasen/ Accessed: 2018-05-02. Archived at: http://www.webcitation.org/6z7kU40qb
49. Röös, Elin. *Mat-klimat-listan*. Rapport 077 (2014). ISSN 1654-9406. 2014.
50. U&We. Report for Systembolaget, “Miljöutredning Systembolaget- Underlag för Systembolagets framtida miljöarbete” [INTERNET] (2011). http://ansvarsredovisning2015.systembolaget.se/wp-content/uploads/2016/01/miljoutredning_systembolaget.pdf Accessed: 2018-05-02. Archived at: http://www.webcitation.org/6z7nBVpJh
51. Florén, Britta, et al. "Lätt att välja rätt–Klimatdata för medvetna val av livsmedelsråvaror i storkök." SP Rapport 47 (2015). ISSN 0284-5172
52. CleanMetrics™. Food Emission Calculator. [INTERNET]. http://www.foodemissions.com/foodemissions/Calculator.aspx Accessed: 2018-05-02. Archived at http://www.webcitation.org/6z7oaSKae
53. Doublet, Geneviève, and Niels Jungbluth. "Life cycle assessment of drinking Darjeeling tea." *Conventional and organic Darjeeling tea. ESU-services Ltd., Uster* (2010).

54. World Health Organization. "Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus." (1999).
55. Moeller, Suzen M., et al. "Dietary patterns: challenges and opportunities in dietary patterns research: an Experimental Biology workshop, April 1, 2006." *Journal of the American Dietetic Association* 107.7 (2007): 1233-1239.
56. Gorst-Rasmussen, Anders. "tt: Treelet transform with Stata." *The Stata Journal*: 12.1.(2012); 130-146
57. Garrett, Joanne M. "PREDXCAT: Stata module to calculate predicted means, medians, or proportions for nominal X's." (2017).
58. Stata, S. "Release 13. Statistical software." *StataCorp LP, College Station, TX* (2013).
59. Rodríguez-Monforte, Míriam, Gemma Flores-Mateo, and Emília Sánchez. "Dietary patterns and CVD: a systematic review and meta-analysis of observational studies." *British Journal of Nutrition* 114.9 (2015): 1341-1359.
60. Brennan, Sarah F., et al. "Dietary patterns and breast cancer risk: a systematic review and meta-analysis–." *The American journal of clinical nutrition* 91.5 (2010): 1294-1302.
61. Cunha, Diana Barbosa, Renan Moritz Varnier Rodrigues de Almeida, and Rosângela Alves Pereira. "A comparison of three statistical methods applied in the identification of eating patterns." *Cadernos de saude publica* 26.11 (2010): 2138-2148.
62. Thorpe, Maree G., et al. "A comparison of the dietary patterns derived by principal component analysis and cluster analysis in older Australians." *International Journal of Behavioral Nutrition and Physical Activity* 13.1 (2016): 30.
63. Winkvist, Anna, et al. "Longitudinal 10-year changes in dietary intake and associations with cardio-metabolic risk factors in the Northern Sweden Health and Disease Study." *Nutrition journal*16.1 (2017): 20.
64. Konde, Å. B., et al. "Swedish dietary guidelines—risk and benefit management report." *Livsmedelverkets rapportserie* 5 (2015).
65. Imamura, Fumiaki, and Paul F. Jacques. "Invited commentary: dietary pattern analysis." *American journal of epidemiology*173.10 (2011): 1105-1108.
66. Lau, Cathrine, et al. "Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study." *British journal of nutrition* 99.5 (2008): 1089-1098.
67. Hu, Frank B., et al. "Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire–." *The American journal of clinical nutrition* 69.2 (1999): 243-249.
68. Heidemann, C., et al. "A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)—Potsdam Study cohort." *Diabetologia* 48.6 (2005): 1126-1134.
69. Esposito, Katherine, et al. "Prevention of type 2 diabetes by dietary patterns: a systematic review of prospective studies and meta-analysis." *Metabolic syndrome and related disorders* 8.6 (2010): 471-476.
70. Pan, An, et al. "Plain-water intake and risk of type 2 diabetes in young and middle-aged women." *The American journal of clinical nutrition* 95.6 (2012): 1454-1460.
71. Roussel, Ronan, et al. "Low water intake and risk for new-onset hyperglycemia." *Diabetes care* 34.12 (2011): 2551-2554.
72. Naumann, Johannes, et al. "Prevention and Therapy of Type 2 Diabetes—What Is the Potential of Daily Water Intake and Its Mineral Nutrients?." *Nutrients* 9.8 (2017): 914.
73. Liu, Enju, et al. "Cross-sectional association of dietary patterns with insulin-resistant phenotypes among adults without diabetes in the Framingham Offspring Study." *British journal of nutrition*102.4 (2009): 576-583.
74. Ocké, Marga C. "Evaluation of methodologies for assessing the overall diet: dietary quality scores and dietary pattern analysis." *Proceedings of the Nutrition Society* 72.2 (2013): 191-199.
75. Wallén, Anna, Nils Brandt, and Ronald Wennersten. "Does the Swedish consumer's choice of food influence greenhouse gas emissions?." *Environmental Science & Policy* 7.6 (2004): 525-535.
76. Hjelm-Wallén, Lena, and Kjell Larsson. "Sveriges klimatstrategi." *Proposition* 2 (2001): 55
77. Murakami, Kentaro, and M. Barbara E. Livingstone. "Greenhouse gas emissions of self-selected diets in the UK and their association with diet quality: is energy under-reporting a problem?." *Nutrition journal* 17.1 (2018): 27.

78. Vittinghoff, Eric, and Charles E. McCulloch. "Relaxing the rule of ten events per variable in logistic and Cox regression." *American journal of epidemiology* 165.6 (2007): 710-718.
79. Hörnell, Agneta, et al. "Mis-reporting, previous health status and health status of family may seriously bias the association between food patterns and disease." *Nutrition journal* 9.1 (2010): 48.
80. Hyland, John J., et al. "Dietary emissions patterns and their effect on the overall climatic impact of food consumption." *International Journal of Food Science & Technology* 52.12 (2017): 2505-2512.
81. Gorst-Rasmussen, Anders, et al. "Gorst-Rasmussen et al. Respond to “Dietary Pattern Analysis”." *American journal of epidemiology* 173.10 (2011): 1109-1110.
82. Imamura, Fumiaki, et al. "Confounding by dietary patterns of the inverse association between alcohol consumption and type 2 diabetes risk." *American journal of epidemiology* 170.1 (2009): 37-45.
83. Bennette, Caroline, and Andrew Vickers. "Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents." *BMC medical research methodology* 12.1 (2012): 21.

Appendix

***Table S1**: Data from life cycle assessment (LCA) studies[i] expressed as carbon dioxide equivalents (CO2e) per kg of food product*

No	FOOD ITEMS LISTED IN THE DIETVIP	Kg CO_2e/Kg of food product
1	Bregott on bread	8
2	Butter on bread	8
3	Low fat margarine on bread	1.5
4	Margarine on bread	1.5
5	Butter for cooking	8
6	Margarine for cooking	1.5
7	Oil for cooking	1.5
8	Salad dressing with oil	1
9	Cream, crème fraiche, sour cream	4
10	Whole grain crisp bread (e.g. Wasa)	0.8
11	Whole grain soft bread	0.8
12	White (soft) and thin crisp bread	0.8
13	Coffee rolls/buns, rusk	0.8
14	Cheese, 28% fat (e.g. Grevè)	8
15	Cheese, 10-17% fat (e.g. Drabant)	8
16	Sausage or liver pàté on bread	7
17	Meat on bread	16[ii]
18	Whole wheat, oat, rye or barley porridge	0.6
19	Rosehip, sweet syrup soup, etc	0.5
20	Sour milk, yoghurt, 3 % fat	1
21	Sour milk, yoghurt, low fat	1
22	Fiber cereals (e.g. muesli, Granola)	0.6
23	Corn flakes	0.6
24	Berries, fresh or frozen	0.2
25	Apple, pear, peach, orange	0.6
26	Banan	1.8
27	Root vegetables, carrot	0.2
28	Tomato, cucumber	1.4
29	Lettuce, lettuce cabbage, white cabbage, spinach, borecole, broccoli	1
30	Boiled or baked potato	0.1
31	Fried potato, pommes frites	0.1
32	Rice	2
33	Pasta	0.8
34	Brown beans, pea soup, and other beans	0.7
35	Pancake, waffle, Swedish (potato) dumpling	0.8[iii]
36	Pizza	5.3[iv]
37	Minced meat dishes	16
38	Meat stew	16[v]
39	Steak, chop etc.	26
40	Bacon	6
41	Saucage as dish	7
42	Hamburger	26
43	White meat (poultry)	3
44	Lean fish (e.g. perch, bass, cod)	3
45	Fatty fish (e.g. (Baltic) herring, whitefish, salmon)	3
46	Salty fish	3
47	Smoked fish/meat	3
48	Ice cream	2
49	Sweets (chocolate, candy)	2
50	Sugar, honey, marmelade, jam	3
51	Cookies, pastry	2[vi]
52	Chips, popcorn, salta nötter, mm	2

53	Low fat milk (0,5%)	1
54	Milk, sour milk (1,5 %)	1
55	Milk (3%)	1
56	Sodas, softdrinks, juice	0.3
57	Brewed (filtered) coffee	3/dry mass
58	Boild coffee	3/dry mass
59	Tea	3/dry mass
60	Light beer	0.49[vii]
61	Medium beer	0.49[vii]
62	Strong beer	0.49[vii]
63	Wine	1.05[vii]
64	Liquor, spirits	1.84
65	Water	0
66	Egg, egg dishes, omelet	2

[i] Values are obtained from LCA data reported by Elin Röös in the 2014 report "Mat-klimat-listan version 1.1", Swedish University of Agricultural Sciences

[ii] Values for "meat on bread" were calculated based on assumption that 50% of consumption coming from beef and another 50% coming from pork

[iii] For pancakes and waffles recipe for plain dishes was used with milk to flour ratio of 75%

[iv] Values for pizza were calculated based on recipe for the most popular Swedish pizza, Kebab pizza as obtained from the Swedish food database The National Food Agency food database, version 2017-12-15. Five ingredients: flour, cheese, tomato sauce, beef meat and dressings were used for the calculation

[v] Value for meat stew was calculated for beef meat cubes in the recipe ratio of 60%

[vi] For cookies and pastries, a recipe with 55% of sugar and 45% of flour was used to calculate cf values based on: Hoolohan, C., Berners-Lee, M., McKinstry-West, J., & Hewitt, C. N. (2013). Mitigating the greenhouse gas emissions embodied in food through realistic consumer choices. Energy Policy, 63, 1065-1074.

[vii] For alcoholic drinks values were obtained from the 2011 U&We report for Systembolaget, "Miljöutredning Systembolaget Underlag för Systembolagets framtida miljöarbete"

***Table S2**. Unadjusted Odd Ratios of Pre-diabetic or Diabetic state (1=IFG/IGT/Diabetes, 0=Normoglycemic) and Quintiles of dietary patterns identified by Treelet Transform in women (n=36,770) and men (n=34,463) in the Västerbotten Intervention Program (VIP), 1996-2013*

	Q1	Q2		Q3		Q4		Q5		P_{trend}
	OR	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	
Women										
TC1	Ref.	1.03	0.95—1.11	**0.89**	0.81—0.96	**0.81**	0.75—0.88	**0.86**	0.80—0.94	***<0.0001***
TC2	Ref.	0.97	0.90—1.06	0.92	0.85—1.00	0.94	0.86—1.02	0.93	0.86—1.01	*0.061*
TC3	Ref.	0.98	0.90—1.06	0.97	0.89—1.05	0.96	0.88—1.04	0.93	0.85—1.01	*0.075*
TC4	Ref.	**0.87**	0.80—0.95	**0.84**	0.78—0.92	**0.82**	0.76—0.89	**0.85**	0.78—0.92	***<0.0001***
Men										
TC1	Ref.	**0.90**	0.83—0.98	**0.83**	0.76—0.90	**0.81**	0.75—0.88	**0.76**	0.70—0.83	***<0.0001***
TC2	Ref.	**0.86**	0.80—0.93	**0.82**	0.75—0.88	**0.69**	0.63—0.75	**0.71**	0.66—0.77	***<0.0001***
TC3	Ref.	**0.80**	0.74—0.88	**0.80**	0.74—0.88	**0.88**	0.82—0.96	**0.89**	0.82—0.97	*0.126*
TC4	Ref.	1.11	1.03—1.21	0.99	0.92—1.09	1.03	0.95—1.12	1.09	1.00—1.19	*0.327*
TC5	Ref.	1.04	0.96—1.14	**1.21**	1.11—1.32	**1.32**	1.21—1.44	**1.40**	1.29—1.53	***<0.0001***

***Table S3**. Unadjusted Odd Ratios of Diabetes (1=Diabetes, 0=Normoglycemic) and Quintiles of dietary patterns identified by Treelet Transform for women (n=30,526) and men (n=28,089) in the Västerbotten Intervention Programme (VIP), 1996-2013*

	Q1	Q2		Q3		Q4		Q5		P_{trend}
	OR	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	
Women										
TC1	Ref.	1.15	0.87—1.51	0.81	0.61—1.09	**0.73**	0.54—0.98	**0.73**	0.54—0.99	***0.001***
TC2	Ref.	0.98	0.73—1.32	0.87	0.64—1.18	1.07	0.80—1.42	0.92	0.68—1.24	*0.806*
TC3	Ref.	0.87	0.65—1.15	1.06	0.81—1.40	**0.66**	0.49—0.90	**0.73**	0.54—0.98	***0.010***
TC4	Ref.	**0.72**	0.55—0.95	**0.56**	0.42—0.74	**0.59**	0.44—0.79	**0.60**	0.45—0.79	***<0.0001***
Men										
TC1	Ref.	0.84	0.65—1.10	0.89	0.69—1.15	0.71	0.54—0.93	0.82	0.63—1.07	*0.062*
TC2	Ref.	0.80	0.63—1.03	**0.59**	0.46—0.77	**0.59**	0.46—0.77	**0.55**	0.42—0.72	***<0.0001***
TC3	Ref.	0.78	0.59—1.02	0.77	0.59—1.02	1.00	0.78—1.30	0.89	0.68—1.16	*0.941*
TC4	Ref.	1.05	0.81—1.38	0.92	0.70—1.21	0.99	0.76—1.30	1.06	0.81—1.39	*0.824*
TC 5	Ref.	1.08	0.81—1.45	**1.40**	1.06—1.86	**1.32**	0.99—1.75	**1.57**	1.19—2.06	***<0.0001***

***Table S4**. Crude and adjusted[†] proportion of "prediabetes or diabetes" and "diabetes" across quintiles of GHGe patterns in men*

TC_1 for carbon footprint in men	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.224	0.196	0.182	*<0.0001*
Adjusted proportion	0.190	0.177	0.174	*0.3189*
Diabetes				
Crude proportion	0.022	0.019	0.018	*0.3967*
Adjusted proportion	0.013	0.012	0.013	*0.7143*

TC_2 for carbon footprint in men	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.245	0.199	0.180	*<0.0001*
Adjusted proportion	0.184	0.185	0.187	*0.1155*
Diabetes				
Crude proportion	0.025	0.016	0.013	*<0.0001*
Adjusted proportion	0.011	0.011	0.012	*0.4159*

TC_3 for carbon footprint in men	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.223	0.188	0.203	*<0.0001*
Adjusted proportion	0.193	0.175	0.175	*0.0713*
Diabetes				
Crude proportion	0.021	0.017	0.019	*0.1460*
Adjusted proportion	0.013	0.011	0.012	*0.6003*

TC_4 for carbon footprint in men	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.197	0.200	0.198	*0.5600*
Adjusted proportion	0.190	0.178	0.176	*0.2452*
Diabetes				
Crude proportion	0.020	0.017	0.019	*0.4978*
Adjusted proportion	0.013	0.010	0.012	*0.3170*

TC_5 for carbon footprint in men	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.217	0.206	0.187	*<0.0001*
Adjusted proportion	0.189	0.180	0.172	*0.0351*
Diabetes				
Crude proportion	0.022	0.018	0.017	*0.1154*
Adjusted proportion	0.012	0.011	0.012	*0.8927*

***Table S5**. Crude & adjusted[†] proportion of "prediabetes or diabetes" and "diabetes" across quintiles of GHGe patterns in women*

TC_1 for carbon footprint in women	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.189	0.180	0.179	*0.0219*
Adjusted proportion	0.155	0.161	0.178	*0.0614*
Diabetes				
Crude proportion	0.018	0.014	0.013	*0.0111*
Adjusted proportion	0.010	0.008	0.007	*0.1083*

TC_2 for carbon footprint in women	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.206	0.175	0.166	*<0.0001*
Adjusted proportion	0.165	0.161	0.169	*0.8908*
Diabetes				
Crude proportion	0.018	0.014	0.013	*0.0118*
Adjusted proportion	0.009	0.008	0.009	*0.8051*

TC_3 for carbon footprint in women	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.167	0.179	0.196	*<0.0001*
Adjusted proportion	0.161	0.160	0.173	*0.5192*
Diabetes				
Crude proportion	0.011	0.016	0.015	*0.0346*
Adjusted proportion	0.009	0.009	0.007	*0.6358*

TC_4 for carbon footprint in women	Q_1	Q_3	Q_5	p-value
IFG or IGT or Diabetes				
Crude proportion	0.202	0.176	0.177	*<0.0001*
Adjusted proportion	0.162	0.169	0.168	*0.5604*
Diabetes				
Crude proportion	0.021	0.012	0.013	*0.0001*
Adjusted proportion	0.010	0.008	0.008	*0.5310*

[†] adjusted for: total energy intake, physical activity, age group, education level, smoking status, alcohol intake, fiber intake, component scores of other dietary patterns, body mass index categories, medication for hypertension, medication for dyslipidemia [as stated in methodology section].

***Table S6**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of cf patterns in female participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables*

	TC_1 for carbon footprint in women				TC_2 for carbon footprint in women			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	1.90 ± 0.39	2.54 ± 0.49	3.76 ± 1.11	<0.001	2.35 ± 0.86	2.63 ± 0.82	3.11 ± 1.00	<0.001
Energy intake (Kcal/d)	1230.24 ± 252.67	1453.71 ± 309.99	1847.03 ± 437.01	<0.001	1269.30 ± 295.56	1449.34 ± 327.97	1820.88 ± 421.37	<0.001
Age group				<0.001				<0.001
35-44 years	25.1%	36.0%	43.9%		14.9%	35.6%	55.4%	
45-54 years	30.4%	31.9%	32.9%		29.8%	34.1%	29.1%	
55-64 years	44.4%	32.1%	23.2%		55.3%	30.3%	15.5%	
BMI group				<0.001				<0.001
underweight	1.4%	1.1%	1.0%		0.9%	0.8%	1.6%	
normal	52.7%	49.9%	47.2%		48.4%	51.6%	52.1%	
overweight	34.1%	33.2%	31.9%		35.4%	32.6%	29.4%	
obese	11.9%	15.9%	19.9%		15.2%	15.0%	17.0%	
Alcohol intake (0 or >20g/d)	11.8%	8.7%	9.6%	<0.001	9.0%	8.6%	12.0%	<0.001
Fiber intake (<15g/1000Kcal)	75.2%	79.1%	82.6%	<0.001	56.1%	82.6%	95.7%	<0.001
Education level				<0.001				<0.001
≤9 years	43.2%	29.2%	19.5%		36.8%	29.3%	24.9%	
high school	29.1%	33.9%	33.6%		26.1%	32.6%	38.7%	
higher education	27.7%	36.9%	46.9%		37.1%	38.1%	36.4%	
Physical activity				<0.001				<0.001
inactive	20.2%	17.0%	14.2%		16.2%	16.2%	17.6%	
moderately inactive	33.1%	28.3%	25.3%		28.7%	28.7%	30.8%	
moderately active	26.4%	28.2%	26.3%		25.6%	27.8%	27.9%	
active	20.3%	26.5%	34.3%		29.4%	27.4%	23.7%	
Smoking status				<0.001				<0.001
non-smoker	47.3%	50.3%	54.5%		45.8%	51.3%	54.3%	
1-4 cigarette	3.3%	2.6%	2.2%		3.0%	2.6%	2.5%	
5-14 cigarette	12.5%	9.2%	6.5%		9.9%	9.0%	8.4%	
15-24 cigarette	3.1%	2.5%	1.7%		2.3%	2.3%	3.2%	
25+ cigarette	0.1%	0.1%	0.1%		0.1%	<1%	0.2%	
occasional smoker	2.5%	2.6%	2.7%		2.4%	2.9%	2.9%	
former occasional smoker	9.6%	10.2%	11.1%		12.0%	10.1%	8.8%	
former smoker	21.6%	22.6%	21.0%		24.5%	21.8%	19.7%	
Medication for hypertension	15.2%	14.3%	12.9%	0.001	19.6%	13.5%	9.1%	<0.001
Medication for dyslipidemia	2.9%	3.1%	3.0%	0.74	5.2%	2.9%	1.2%	<0.001
IFG or IGT or Diabetes	18.9%	18.0%	17.9%	0.022	20.6%	17.5%	16.6%	<0.001
Diabetes	1.8%	1.4%	1.3%	0.011	1.8%	1.4%	1.3%	0.011

***Table S6**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of cf patterns in female participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables (continued)*

	TC_3 for carbon footprint in women				TC_4 for carbon footprint in women			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	2.46 ± 0.87	2.58 ± 0.85	3.10 ± 0.98	<0.001	2.64 ± 0.97	2.65 ± 0.87	2.77 ± 0.91	<0.001
Energy intake (Kcal/d)	1251.80 ± 278.80	1446.29 ± 326.09	1848.73 ± 411.72	<0.001	1509.69 ± 410.58	1463.66 ± 376.37	1516.36 ± 391.34	<0.001
Age group				<0.001				<0.001
35-44 years	49.1%	33.2%	26.7%		35.6%	36.1%	31.3%	
45-54 years	32.2%	32.6%	30.7%		27.4%	34.4%	33.4%	
55-64 years	18.8%	34.1%	42.5%		36.9%	29.5%	35.3%	
BMI group				<0.001				<0.001
underweight	1.4%	1.0%	1.0%		1.5%	0.9%	0.8%	
normal	54.4%	49.9%	47.4%		43.2%	52.6%	56.6%	
overweight	30.4%	34.5%	34.1%		33.1%	32.8%	31.5%	
obese	13.8%	14.6%	17.5%		22.2%	13.7%	11.0%	
Alcohol intake (0 or >20g/d)	8.8%	9.3%	11.1%	<0.001	43.7%	0.0%	1.4%	<0.001
Fiber intake (<15g/1000Kcal)	95.5%	80.5%	59.4%	<0.001	76.7%	76.7%	84.9%	<0.001
Education level				<0.001				<0.001
≤9 years	23.2%	30.3%	36.5%		37.5%	25.8%	24.6%	
high school	38.7%	31.5%	28.3%		33.3%	33.9%	26.8%	
higher education	38.0%	38.2%	35.2%		29.1%	40.3%	48.6%	
Physical activity				<0.001				<0.001
inactive	20.5%	16.5%	14.3%		17.3%	16.3%	17.7%	
moderately inactive	30.1%	29.6%	26.8%		30.0%	27.9%	30.4%	
moderately active	26.7%	26.9%	27.2%		26.9%	27.4%	25.5%	
active	22.7%	27.0%	31.6%		25.8%	28.5%	26.4%	
Smoking status				<0.001				<0.001
non-smoker	47.5%	50.9%	54.4%		64.6%	48.1%	42.7%	
1-4 cigarette	3.0%	2.7%	2.5%		1.9%	2.6%	3.5%	
5-14 cigarette	11.1%	9.0%	7.0%		7.8%	9.3%	8.2%	
15-24 cigarette	3.8%	2.2%	1.4%		2.3%	2.2%	2.7%	
25+ cigarette	0.1%	0.1%	0.1%		0.1%	<1%	0.1%	
occasional smoker	3.4%	2.7%	2.2%		0.9%	3.0%	4.0%	
former occasional smoker	10.0%	10.9%	10.3%		6.1%	11.6%	13.4%	
former smoker	21.3%	21.6%	22.1%		16.2%	23.1%	25.5%	
Medication for hypertension	10.9%	14.2%	16.3%	<0.001	16.4%	13.6%	13.1%	<0.001
Medication for dyslipidemia	2.1%	3.5%	3.8%	<0.001	3.9%	3.0%	2.8%	<0.001
IFG or IGT or Diabetes	16.7%	17.9%	19.6%	<0.001	20.2%	17.6%	17.7%	<0.001
Diabetes	1.1%	1.6%	1.5%	0.042	2.1%	1.2%	1.3%	<0.001

***Table S7**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of cf patterns in male participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables*

	TC_1 for carbon footprint in men				TC_2 for carbon footprint in men			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	2.40 ± 0.55	3.19 ± 0.67	4.80 ± 1.50	<0.001	2.91 ± 1.12	3.31 ± 1.03	4.0 ± 1.35	<0.001
Energy intake (Kcal/d)	1596.88 ± 340.14	1935.21 ± 418.16	2526.63 ± 583.59	<0.001	1727.80 ± 442.63	1932.27 ± 473.30	2401.79 ± 579.47	<0.001
Age group				<0.001				<0.001
35-44 years	29.3%	33.6%	41.6%		14.8%	36.1%	51.7%	
45-54 years	28.0%	32.2%	33.8%		27.2%	33.8%	29.7%	
55-64 years	42.7%	34.3%	24.6%		57.9%	30.1%	18.6%	
BMI group				<0.001				<0.001
underweight	0.3%	0.2%	0.1%		0.3%	0.2%	0.2%	
normal	38.3%	34.2%	31.4%		31.2%	34.0%	37.7%	
overweight	49.6%	50.1%	48.5%		51.1%	50.3%	47.0%	
obese	11.8%	15.5%	19.9%		17.4%	15.5%	15.1%	
Alcohol intake (0 or >20g/d)	8.1%	7.0%	10.1%	<0.001	8.5%	7.7%	10.5%	<0.001
Fiber intake (<15g/1000Kcal)	96.2%	94.9%	94.3%	<0.001	84.8%	97.4%	99.4%	<0.001
Education level				<0.001				<0.001
≤9 years	49.1%	35.5%	26.6%		46.1%	34.0%	30.3%	
high school	34.2%	37.7%	38.5%		28.2%	38.7%	43.1%	
higher education	16.7%	26.9%	34.9%		25.7%	27.3%	26.7%	
Physical activity				<0.001				<0.001
inactive	19.8%	17.5%	15.0%		17.9%	17.9%	17.5%	
moderately inactive	31.4%	29.5%	25.0%		28.6%	28.5%	28.6%	
moderately active	30.4%	28.8%	28.1%		27.3%	29.2%	30.8%	
active	18.4%	24.2%	31.9%		26.1%	24.4%	23.2%	
Smoking status				<0.001				<0.001
non-smoker	46.5%	52.9%	57.0%		44.30%	54.1%	59.0%	
1-4 cigarette	3.0%	2.4%	2.5%		2.8%	2.5%	2.7%	
5-14 cigarette	7.4%	5.0%	4.4%		6.0%	5.0%	5.1%	
15-24 cigarette	4.9%	2.7%	2.3%		3.5%	2.8%	3.2%	
25+ cigarette	0.5%	0.3%	0.3%		0.5%	0.2%	0.4%	
occasional smoker	3.1%	3.7%	3.0%		2.5%	3.8%	3.3%	
former occasional smoker	8.8%	10.1%	10.8%		11.0%	10.0%	8.4%	
former smoker	25.9%	23.1%	19.7%		29.4%	21.5%	17.9%	
Medication for hypertension	14.7%	15.4%	13.4%	0.010	22.4%	13.2%	8.9%	<0.001
Medication for dyslipidemia	5.5%	6.0%	5.9%	0.81	10.4%	5.1%	3.0%	<0.001
IFG or IGT or Diabetes	22.4%	19.6%	18.2%	<0.001	24.5%	19.9%	18.0%	<0.001
Diabetes	2.2%	1.9%	1.8%	0.38	2.5%	1.6%	1.3%	<0.001

***Table S7**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of cf patterns in male participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables (continued)*

	TC_3 for carbon footprint in men				TC_4 for carbon footprint in men			
	Q_1	Q_3	Q_5	p-value	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	3.27 ± 1.30	3.37 ± 1.15	3.60 ± 1.27	<0.001	4.15 ± 1.59	3.11 ± 1.0	3.28 ± 1.05	<0.001
Energy intake (Kcal/d)	1986.75 ± 560.58	1959.25 ± 529.61	2082.27 ± 554.36	<0.001	2020.49 ± 574.53	1895.72 ± 502.82	2206.50 ± 581.46	<0.001
Age group				<0.001				<0.001
35-44 years	33.4%	40.2%	27.6%		47.1%	33.2%	26.7%	
45-54 years	30.5%	30.0%	31.6%		28.7%	30.8%	34.1%	
55-64 years	36.1%	29.7%	40.8%		24.2%	36.0%	39.2%	
BMI group				<0.001				<0.001
underweight	0.5%	0.2%	0.2%		0.2%	0.2%	0.2%	
normal	33.8%	34.4%	35.8%		30.2%	35.8%	36.7%	
overweight	46.3%	49.3%	51.4%		48.9%	49.7%	49.1%	
obese	19.4%	16.2%	12.6%		20.6%	14.3%	14.0%	
Alcohol intake (0 or >20g/d)	25.2%	0.0%	16.1%	<0.001	8.8%	7.2%	9.7%	<0.001
Fiber intake (<15g/1000Kcal)	93.0%	95.4%	96.5%	<0.001	99.8%	98.5%	81.7%	<0.001
Education level				<0.001				<0.001
≤9 years	44.3%	35.8%	28.5%		33.0%	39.1%	33.6%	
high school	35.0%	39.8%	32.5%		43.4%	35.6%	33.0%	
higher education	20.6%	24.3%	39.0%		23.6%	25.3%	33.5%	
Physical activity				<0.001				<0.001
inactive	17.8%	17.1%	19.5%		18.6%	18.6%	14.7%	
moderately inactive	27.7%	27.7%	30.8%		28.3%	28.8%	26.0%	
moderately active	29.7%	29.9%	26.5%		30.9%	29.7%	27.9%	
active	24.8%	25.3%	23.2%		22.2%	22.9%	31.3%	
Smoking status				<0.001				<0.001
non-smoker	60.6%	52.4%	45.7%		50.4%	52.8%	54.9%	
1-4 cigarette	2.0%	2.7%	3.2%		3.1%	2.6%	2.3%	
5-14 cigarette	5.7%	5.5%	4.9%		6.9%	4.7%	3.6%	
15-24 cigarette	3.1%	3.3%	3.3%		5.1%	2.9%	1.2%	
25+ cigarette	0.4%	0.2%	0.5%		0.6%	0.4%	0.2%	
occasional smoker	1.4%	3.7%	4.1%		4.3%	3.6%	2.6%	
former occasional smoker	7.3%	10.3%	11.3%		9.5%	9.7%	11.0%	
former smoker	19.5%	21.9%	27.0%		20.2%	23.3%	24.2%	
Medication for hypertension	16.5%	14.1%	14.6%	<0.001	11.7%	14.4%	16.2%	<0.001
Medication for dyslipidemia	6.7%	6.1%	6.1%	<0.001	4.4%	5.8%	7.4%	<0.001
IFG or IGT or Diabetes	22.3%	18.8%	20.3%	<0.001	19.7%	20.0%	19.8%	0.56
Diabetes	2.1%	1.7%	1.9%	0.15	2.0%	1.7%	1.9%	0.50

***Table S7**. Unadjusted distribution of estimated emission, energy intake and background variables across quintiles of cf patterns in male participants in the VIP, between 1996-2013. Data values are presented as mean values ± SDs for continuous variables and as frequency (%) for categorical variables (continued)*

	TC_5 for carbon footprint in men			
	Q_1	Q_3	Q_5	p-value
Estimated GHGe (Kg CO_2e/d)	3.12 ± 1.17	3.27 ± 1.14	3.84 ± 1.31	<0.001
Energy intake (Kcal/d)	1806.99 ± 465.12	1945.19 ± 500.30	2362.24 ± 583.44	<0.001
Age group				<0.001
35-44 years	30.2%	31.3%	36.9%	
45-54 years	32.2%	32.1%	30.7%	
55-64 years	37.6%	36.6%	32.4%	
BMI group				<0.001
underweight	0.2%	0.2%	0.3%	
normal	30.4%	34.7%	39.7%	
overweight	52.0%	49.4%	45.7%	
obese	17.4%	15.7%	14.2%	
Alcohol intake (0 or >20g/d)	8.3%	7.7%	9.0%	0.11
Fiber intake (<15g/1000Kcal)	89.1%	95.8%	98.5%	<0.001
Education level				<0.001
≤9 years	35.7%	39.2%	38.8%	
high school	36.0%	35.7%	37.6%	
higher education	28.3%	25.0%	23.6%	
Physical activity				<0.001
inactive	18.2%	18.2%	15.7%	
moderately inactive	29.1%	29.6%	26.6%	
moderately active	27.3%	29.2%	31.8%	
active	25.3%	23.0%	25.9%	
Smoking status				<0.001
non-smoker	50.6%	52.0%	54.1%	
1-4 cigarette	2.5%	2.5%	2.8%	
5-14 cigarette	4.5%	5.7%	5.7%	
15-24 cigarette	2.6%	3.1%	4.2%	
25+ cigarette	0.2%	0.3%	0.5%	
occasional smoker	3.2%	3.3%	3.0%	
former occasional smoker	9.7%	9.8%	9.8%	
former smoker	26.6%	23.3%	20.0%	
Medication for hypertension	18.2%	14.6%	12.3%	<0.001
Medication for dyslipidemia	9.0%	5.9%	3.8%	<0.001
IFG or IGT or Diabetes	21.7%	20.6%	18.7%	<0.001
Diabetes	2.2%	1.8%	1.7%	0.11