

RESEARCH

Open Access



# Use of dietary indices to control for diet in human gut microbiota studies

Ruth C. E. Bowyer<sup>1</sup>, Matthew A. Jackson<sup>1</sup>, Tess Pallister<sup>1,2</sup>, Jane Skinner<sup>3</sup>, Tim D. Spector<sup>1</sup>, Ailsa A. Welch<sup>3</sup> and Claire J. Steves<sup>1,4\*</sup>

## Abstract

**Background:** Environmental factors have a large influence on the composition of the human gut microbiota. One of the most influential and well-studied is host diet. To assess and interpret the impact of non-dietary factors on the gut microbiota, we endeavoured to determine the most appropriate method to summarise community variation attributable to dietary effects. Dietary habits are multidimensional with internal correlations. This complexity can be simplified by using dietary indices that quantify dietary variance in a single measure and offer a means of controlling for diet in microbiota studies. However, to date, the applicability of different dietary indices to gut microbiota studies has not been assessed. Here, we use food frequency questionnaire (FFQ) data from members of the TwinsUK cohort to create three different dietary measures applicable in western-diet populations: The Healthy Eating Index (HEI), the Mediterranean Diet Score (MDS) and the Healthy Food Diversity index (HFD-Index). We validate and compare these three indices to determine which best summarises dietary influences on gut microbiota composition.

**Results:** All three indices were independently validated using established measures of health, and all were significantly associated with microbiota measures; the HEI had the highest *t* values in models of alpha diversity measures, and had the highest number of associations with microbial taxa. Beta diversity analyses showed the HEI explained the greatest variance of microbiota composition. In paired tests between twins discordant for dietary index score, the HEI was associated with the greatest variation of taxa and twin dissimilarity.

**Conclusions:** We find that the HEI explains the most variance in, and has the strongest association with, gut microbiota composition in a western (UK) population, suggesting that it may be the best summary measure to capture gut microbiota variance attributable to habitual diet in comparable populations.

**Keywords:** Microbiome, Microbiota, Dietary Index, Dietary covariate, Human microbiota, Food frequency questionnaire, FFQ, Healthy Eating Index, HEI, Mediterranean Dietary Score, MDS, Healthy Food Diversity Index, HFD-Index

## Background

The composition of the gut microbiota is associated with various aspects of human health and by many is considered a new clinical target [1]. Genetic influences are thought to be low, with environmental factors being the primary drivers of variation [2, 3]. Research has focused on host-mediated environmental factors such as xenobiotic exposure, antibiotic use and, in particular, diet,

where multiple studies have indicated associations of long-term diet with the microbiota [4–6]. For example, non-digestible fermentable dietary carbohydrates, short-chain fatty acid ratios and dietary protein and fat can modulate bacterial abundance [7–11]. However, the extent to which clinical interventions or more distal factors, such as socio-economics and geo-physical factors influence the microbiota are emerging questions [12–14]. Selecting a dietary measure which encapsulates the variance in the microbiota attributable to diet is a useful goal which enables adjustment for diet in many studies. However, currently there is no standard approach to quantification of dietary data in microbiota studies.

\* Correspondence: [claire.j.steves@kcl.ac.uk](mailto:claire.j.steves@kcl.ac.uk)

<sup>1</sup>The Department of Twin Research, Kings College London, 3-4th Floor South Wing Block D, St Thomas' Hospital, Westminster Bridge Road, London SE1 7EH, UK

<sup>4</sup>Clinical Age Research Unit, Kings College Hospital Foundation Trust, London, UK

Full list of author information is available at the end of the article



Diet is a complex, multi-faceted phenotype that is often summarised using dietary indices to simplify analyses [15]. Dietary indices are nutritionally derived indices based on levels of (often differently defined) healthy consumption of nutrients or food groups. Analysing diet with the focus on patterns rather than individual dietary constituents is advantageous because dietary constituents are consumed together and often correlate with one another [16]. Dietary indices therefore provide a means to capture the overall dietary pattern of an individual or population in a single measure, allowing adequate adjustment for diet without saturation of models by the high dimensionality of dietary data. Dietary indices tend to assess diet quality based broadly on one of three categories; consumption measured against dietary guidelines, recommend foods, and dietary variety [17]. Indices within this analysis were selected to fall broadly into one of these three categories and because they were not defined in relation to a specific disease.

#### Dietary indices

##### *Healthy Eating Index (HEI)*

The Healthy Eating Index (HEI) 2010 is a dietary index developed by the United States Department of Agriculture (USDA) as a means to assess diet measured as compliance to US Dietary Guidelines for Americans [15]. Designed to capture diet quality from 24-h food recalls and FFQ data, the HEI is comprised of 12 calorie-adjusted components representing 'adequacy' components, scored to reflect the extent an individual meets the recommended consumption level for that group, and 'moderation' components, where maximum scores are awarded when consumption falls below a lower threshold. The HEI is scored from 0 to 100; the higher an individual's score therefore, the healthier their diet is considered to be. The HEI was selected for this analysis because it is readily applicable to FFQ data [18]; it contains relative weighted measures for each group; and because it uses set thresholds (i.e. rather than those based on study population averages).

##### *Mediterranean Dietary Score (MDS)*

Mediterranean diets are associated with lower rates of common chronic diseases. They are characterised by high intakes of whole grain, vegetables, legumes, fruit, unsaturated lipids and fish; low to medium intakes of saturated lipids, meat and dairy, and modest alcohol consumption [19]. The Mediterranean Dietary Score (MDS), scored from 0 to 10, is considered here as an index based on study population averages; due to its increasing popularity as a measure of dietary health [20]; and because of its straightforward method of grouping foods. Here, we use methodology developed and evaluated for use in non-Mediterranean countries [19].

##### *Healthy Food Diversity index (HFD-index)*

Indices that capture dietary diversity may offer researchers a fast and effective way of assessing dietary quality, based on suggestive evidence that a more diverse diet may be associated with better health outcomes [21]. In addition, we hypothesised that a wider variety of foods may result in a wider variety of ecological niches for microbes. The Healthy Food Diversity index (HFD-index) scores between 0 and  $1 - 1/\text{number of individuals}$  (0.9998 for this study), where a higher value indicates a more diverse diet. The HFD-index was selected for this study as it considers diversity of food in conjunction with using a weighted health value to circumnavigate many of the traditional problems of measures of dietary diversity [22], and has been used in a previous microbiota study as a dietary covariate [23].

In this analysis, we first validate each index as a measure of a healthy diet within the TwinsUK cohort, and then assess each index's association with measures of gut microbiota composition. Our aim is to determine the optimal summary measure of diet-based variation in gut microbiota composition for use as a covariate in future analyses.

#### Results

All dietary indices were validated within the TwinsUK cohort, with results suggesting all three capture diet successfully. Microbiota associations were observed with all three indices, with the greatest number being associated with the HEI.

##### **Index construction and validation**

Index scores created from data of 5047 individuals were used to assess index validity (Table 1). None of the indices achieved minimum or maximum scores possible in their 1st and 99th percentile, as expected given the real-world nature of the data (Additional file 1: Table S1). The range of all of the indices was wide enough to allow meaningful differences to be detected.

Based on previous research, dietary indices are expected to be predictive of differences between populations known to have differing dietary patterns. In this case, concurrent criterion validation suggests that a dietary index predicts non-smokers, women and older people to have healthier diets than smokers, men and younger people, respectively [18]. All three indices significantly predicted a difference of means for smoking and non-smoking; the HEI and MDS for men and women, and just the MDS for age (Table 2).

Both the HEI and the MDS had a small, significant negative association with BMI; there was no significant association with the HFD-index (Table 3). HEI and MDS had a small but significant negative association with health as captured by the frailty index (where age, zygosity and

**Table 1** Descriptive statistics of validation cohort and microbiota subset

Characteristic (measure)	Validation cohort	Microbiota subset
<i>n</i>	5047	2070
Sex (%female)	91.2	90
Zygoty (% MZ)	56.8	55.9
Ethnicity (% white)*	98.2	98.6
Age (at FFQ) ( $\mu, \sigma^2$ )	58.4 (13.2)	60.5 (11.5)
BMI ( $\mu, \sigma^2$ )*	26.2 (5)	25.9 (4.7)
FI ( $\mu, \sigma^2$ )*	0.2 (0.1)	0.19 (0.1)
HEI ( $\mu, \sigma^2$ )	60 (10.3)	60.4 (10.2)
MDS ( $\mu, \sigma^2$ )	4.6 (1.8)	4.5 (1.8)
HFD-index ( $\mu, \sigma^2$ )	0.2 (0.1)	0.2 (0.1)

Descriptive statistics of cohorts used to validate (validation cohort) three dietary indices and assess association with the microbiota (microbiota subset). Presented also are the means ( $\mu$ ) and standard deviation ( $\sigma^2$ ) of the three indices: the Healthy Eating Index (HEI), the Mediterranean Diet Score (MDS) and Healthy Food Diversity index (HFD-index). Zygoty is presented as % mono-zygotic twins (MZ), age as the date at which the Food Frequency Questionnaire (FFQ) was administered, body mass index by BMI ( $\text{kg}/\text{m}^2$ ), and frailty index (proportion of age-related health-deficits) as FI. \*Data available on 88–90% subjects for these variables

sex were covariates). The frailty index (FI) is the proportion of age-related health deficits reported by subjects from over 30 holistic health domains [24]; the HFD-index exhibited a small positive association with FI suggesting that diversity of food is associated with adverse health (Table 3).

### Microbiota assessment

A subset of 2070 individuals with 16S rRNA gene sequencing gut microbiota data were used to assess the extent the dietary indices were able to explain variance within the cohorts microbial community structure (Table 1). Linear mixed-effects models were used to assess associations between the dietary indices and alpha diversity. All significant associations were small, with the highest  $\beta$  value observed between Shannon diversity and the HEI (Fig. 1, Table 4). The highest  $t$  values came from the HEI, where many were greater than 2, the threshold for indication of good model fit [25] (Additional file 1: Table S3). Both the HEI and MDS were significantly associated with

**Table 2** Concurrent criterion validation of dietary indices

	<i>n</i>	HEI	MDS	HFD-index
Men vs women	443:4604	56.4:60.4***	4.2:4.6***	0.2:0.19 $\perp$
Over 60s vs under 60s	2543:2504	59.8:60.3 $\perp$	4.5:4.6**	0.2:0.2 $\perp$
Smokers vs non-smokers	317:2909	55.9:61***	4.1:4.7***	0.16:0.21***

Three dietary indices, the Healthy Eating Index (HEI), the Mediterranean Diet Score (MDS) and the Healthy Food Diversity index (HFD-index) were assessed for their ability to predict difference of diet of smokers vs non-smokers, over 60s vs under 60s, and men vs women via two sample  $t$  test (HFD-index via Wilcoxon rank sum). Difference in means is displayed for each grouping, with significance thresholds indicated by: \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ,  $\perp$  = non-significant. Results of tests are indicated in Additional file 1: Table S2

**Table 3** Correlation of dietary indices with health measures

	<i>n</i>	HEI	MDS	HFD-index
BMI	4428	$\beta = -0.076$ ***	$\beta = -0.098$ ***	Non-significant
FI	4553	$\beta = -0.12$ ***	$\beta = -0.11$ ***	$\beta = 0.013$ *

Three dietary indices, the Healthy Eating Index (HEI), the adjusted-Mediterranean Diet Score (MDS) and the Healthy Food Diversity index (HFD-index) were assessed for their correlation with two health measures; body mass index (BMI  $\text{kg}/\text{m}^2$ ) and Rockwood's frailty index (FI) [24] via nested linear regression models (adjusting for age, sex and zygoty).  $p$  values:  $p < 0.05$ \*,  $p < 0.001$ \*\*\*

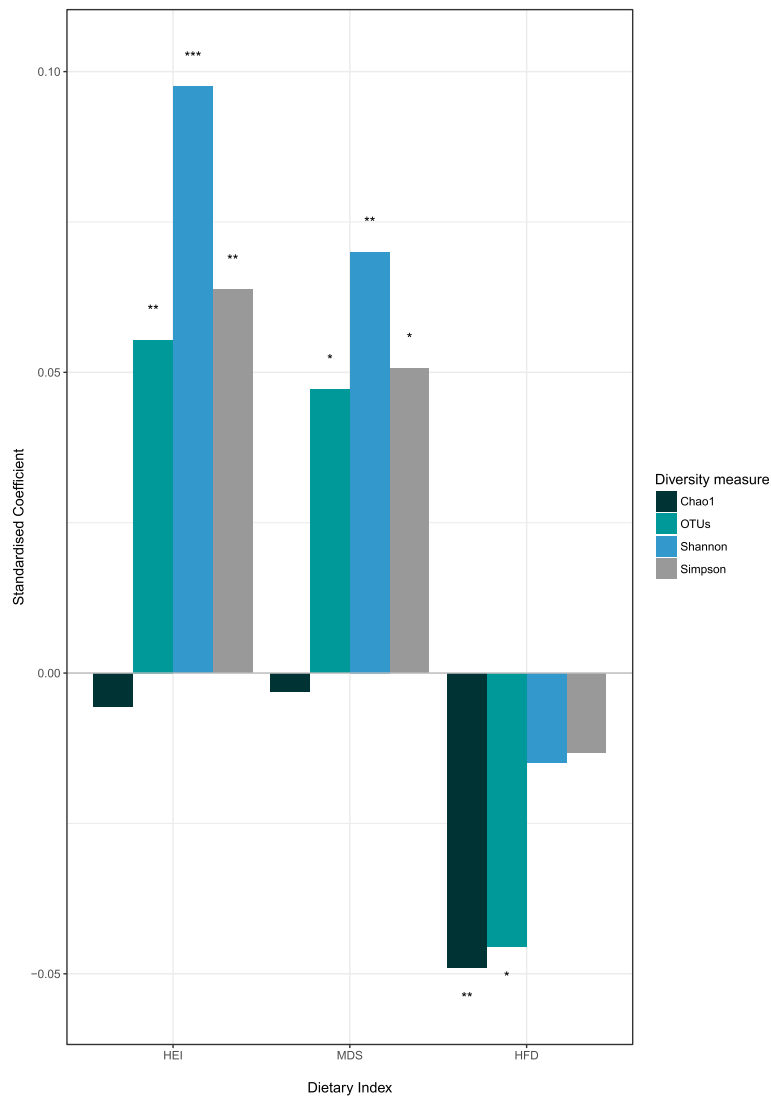
number of OTUs, Shannon and Simpson indices; only the HFD-index was associated with diversity indicator Chao1. Interestingly, all alpha diversity associations with the HFD-index were negative. Comparison of  $t$  values from the HEI, MDS and HFD-index found that the HEI explains more of the variance within the data than the other two indices (Additional file 1: Table S3).

We used hierarchical modelling to investigate the contributions to variance explained by health and diet separately and together (Additional file 1: Table S4–6). Beta coefficients are similar across all models suggesting dietary indices capture alpha diversity variance attributable to diet independent of health deficits.

All three indices exhibited FDR-adjusted associations with individual OTU relative abundances significant at  $q < 0.05$ : the HEI had 167, the MDS had 107, and the HFD-index had 13 (Table 5, Additional file 1: Table S7–13). Both the HEI and MDS exhibited significant negative correlations with *Ruminococcus*, *Lachnospira* and *Actinomyces* (Additional file 1: Table S9–13). The HFD-index also exhibited significant correlations with several *Ruminococcus* and *Lachnospiraceae*; with only one genus-level association assigned to genus Cc115 within the family *Erysipelotrichaceae*.

In linear mixed-effects models, the HEI was significantly associated with axes 1, 2, 4, 8 and 10 from PCoA of unweighted UniFrac distances; the MDS with the first 2 and the highest correlations for both was with axis 2 (HEI:  $\beta = -0.14$ ,  $p < 0.0001$ , MDS:  $\beta = -0.12$ ,  $p < 0.0001$ ) (Additional file 1: Table S14–15). The HFD-index was approaching significance with axis 2 ( $\beta = -0.039$ ,  $p = 0.055$ ) and axis 8 ( $\beta = -0.095$ ,  $p < 0.0001$ ) (Additional file 1: Table S16).

The unique setting of this study within a large twin cohort allowed us to undertake twin paired tests that reduce the variation due to genetic and early-environmental factors. Twins discordant for their dietary index value were assessed using paired Wilcoxon rank-sum tests to replicate OTU associations. We observed that of the 167 HEI-associated OTUs, 71 were nominally significant in difference between “healthy diet” to “less healthy diet” pairs, and 17 were FDR-adjusted significant to  $q < 0.05$  (Fig. 2). Of the 107 OTUs associated with the MDS, 32 were nominally significant and one, an OTU assigned to genus *Coprococcus*, was FDR significant. Of the 13 FDR-



**Fig. 1** Standardised coefficients indicating correlation magnitude from mixed-effects models of three dietary indices (the Healthy Eating Index (HEI), Mediterranean Diet Score (MDS) and the Healthy Food Diversity index (HFD-index) for four measures of microbial alpha diversity; Chao1, observed OTUs, Shannon index and the Simpson index. Only significant results included, *p* values are \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001. Full results, including model AIC and *t* values are in Additional file 1: Table S3. Alpha diversity metrics were rarefied and adjusted for age, sex, gender and technical covariates

**Table 4** Alpha diversity results

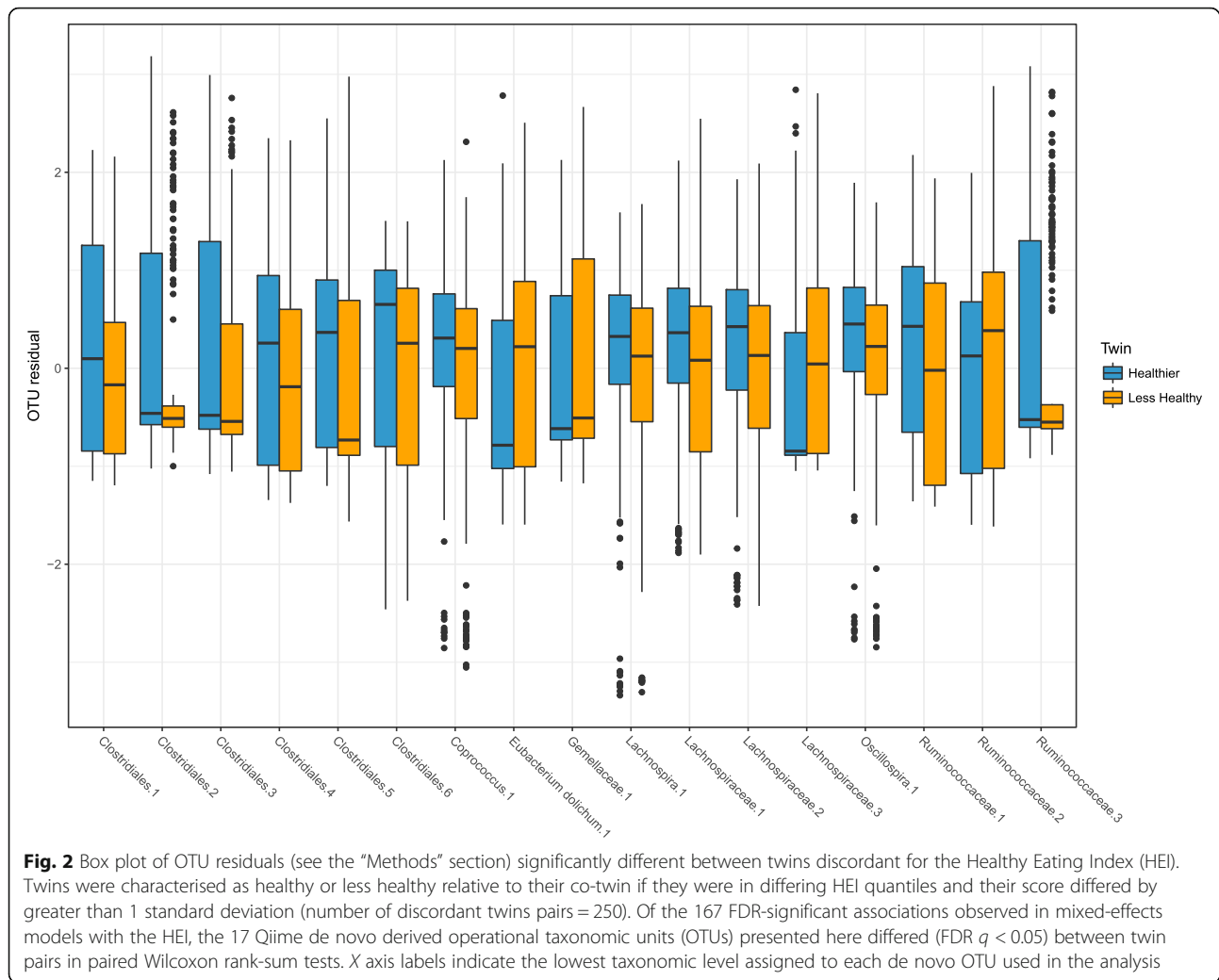
Diversity measure	HEI	MDS	HFD-Index
Chao1	-0.01 $\perp$	> 0.00 $\perp$	- 0.05**
OTUs	0.06**	0.05*	- 0.05*
Shannon	0.1***	0.07***	- 0.01 $\perp$
Simpson	0.06**	0.05*	- 0.01 $\perp$

Standardised coefficients of linear mixed-effects models of three dietary indices, the Healthy Eating Index (HEI), Mediterranean Diet Score (MDS) and the Healthy Food Diversity index (HFD-index), against four measures of alpha diversity (Chao1, Observed number of OTUs (OTUS), Shannon diversity and Simpson's diversity index). Alpha diversity measures were rarefied and adjusted for BMI, sex, age and technical covariates (see the "Methods" section). *p* values: \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001,  $\perp$  = non-significant

**Table 5** Number of taxonomic associations observed with dietary indices

Taxonomic level	HEI	MDS	HFD-index
OTUs	167	107	13
Genus	16	6	1
Phylum	4	0	0

Number of Qiime de novo derived operational taxonomic units (OTUs), genus and phyla significantly associated with three dietary indices, the Healthy Eating Index (HEI), Mediterranean Diet Score (MDS) and the Healthy Food Diversity index (HFD-index). Number of results are those significant post FDR adjustments in linear mixed effects models adjusted for age, BMI, sex and technical microbiota covariates. Full results are included in Additional file 1: Table S7-13



adjusted significant associations with the HFD-index, none were significantly associated with discordant twins. In regression analyses of weighted UniFrac distance between 755 twin pairs against dietary index dissimilarity, adjusted for difference of BMI and technical covariates, no significant associations were observed.

## Discussion

The primary aim in this analysis was to identify a dietary composite which explains variation in the gut microbiota, and therefore might have most utility to capture diet in microbiota studies. In this analysis, three dietary indices were successfully applied to FFQ data derived from the TwinsUK cohort and were assessed for their ability to explain inter-individual variance within the gut microbiota. Our evidence here is suggestive of the HEI being the index of choice.

We made some assumptions in this analysis; that it is the range of healthy diets along these indices that captures the greatest range of difference between microbial

communities; that the dietary index that captures the highest amount of variance with measures of alpha and beta diversity, and the highest number of associations with OTUs is the index of preference. However, we make no assumption that the microbes associated with the higher dietary scores are necessarily the microbes that are the most important for health.

All three dietary indices performed in validation tests and could therefore have specific utility. All could distinguish between smokers and non-smokers; the HEI and MDS differed marginally between women and men; only the MDS could distinguish between young and old, but the magnitude of the effect was minimal. The HEI and MDS were significantly negatively associated with frailty as would be expected; the frailer a person, the less healthy their diet [26], further confirming validity of the HEI and MDS as a measure of healthy eating. Associations with health measures (BMI and frailty) were small, as expected due to the large number of factors influencing health [18, 24]. The positive HFD-index association with frailty, although small,



was in an unexpected direction and is difficult to interpret. This may reflect the fact that the HFD-index was not calorie adjusted, whereas the other indices take total energy content into account. One concurrent criterion proposed by Guenther was that the Healthy Eating Index was improved in older adults compared with younger adults [18]. Our sample detected a small difference between age groups for the MDS, but no difference for the HEI. As the HEI is population independent, this may be a consequence of our study population demographics (older, white, middle-class women) with limited sampling in the younger age groups. The MDS could have succeeded here because it reflects dietary preference of specific food groups relative to the study population mean.

Our focus was to find a means of controlling for as much dietary-influenced microbiota variation as possible. Therefore, as the HEI had the greatest association with the microbiota and explained the most variance and dissimilarity of the data within this cohort, we argue that it can be deemed the most suitable index to use as a dietary covariate. Del Chierico 2014 [20] makes the a priori assumption that there will be compelling evidence for microbiota associations with the MDS based on its positive associations with health outcomes. Indeed, we observed, like others [19], the MDS to have associations with health measures and microbiota. However, the nature of the HEI (comprised of multifaceted components rather than binary variables and with a larger numeric range) means it covers more variation of diet compared with the MDS (with associations based on population medians and comprised of a much smaller numeric range). This may explain its larger capture of microbiota associations within our population.

The HFD-index also exhibited some intriguing associations as an index based on dietary variety; and the different outcomes when compared with the HEI and MDS may suggest some underlying associations driven by diversity of diet. However, its negative associations with alpha diversity are at odds with what might be hypothesised; that a more varied diet creates more ecological niches for a more diverse community assemblage. It is likely that patterns observed here are due to unsuitability of FFQ raw data for this index; many of the values used to create the health value were difficult to ascertain in quantities from the data (e.g. wheat germ oil and soy bean oil). Indeed, the FFQ has been designed to capture intakes of the most frequently consumed food for a population; therefore, inherent in the data is a limit on its ability to capture diversity of diet. Additionally, whilst diet-diversity indices are frequently utilised as indicators of nutrient intake in children and populations from lower-income countries [27, 28] they may be less suited to western diets.

The HEI can be appropriately applied to a wide range of dietary data types; particularly 24 h and 3-day recall diaries,

and therefore offers opportunity as the covariate of choice for a wide range of microbiota studies [15, 18]. OTU associations that differed between twin pairs discordant for the HEI generally followed health-associated patterns previously observed. *Eubacterium dolichum*, associated with a lower (less-healthy) HEI score in the present study, was observed to positively associate with frailty [29] and with a dietary score based on visceral fat mass within this cohort [30]. This finding is in keeping with Murine models showing blooming of related bacteria (*Erysipelotrichi*) in the context of an unhealthy diet [5]; similarly, genus *Oscillospira* (here, associated positively with HEI) has been observed to be reduced in the presence of diseases that involve inflammation and patients with non-alcoholic fatty liver disease [31], and was negatively correlated with BMI differences in a different twin cohort [32]. *Clostridiales* are a polyphyletic group with some notable pathogenic gut species (e.g. *Clostridium difficile*), yet contribute in force to the core microbiota [33]. Their decreased relative abundance has been shown to associate with disease states and here was enriched in twins with a higher HEI score relative to their sibling. Similarly, an increase in *Fusobacteriawas* associated with disease states and was observed here in higher relative abundance in less healthy eaters [34]. *Lachnospiraceae* were less enriched in colorectal cancer patients and again here mostly associated with a higher HEI score [35]. Therefore, this suggests that the HEI is associated with bacterial species in a way that would be expected given its design as a measure of healthy diet, and is applicable as a means of explaining dietary impact on the community composition of the microbiota.

A key consideration in the utilisation of FFQ data is its appropriateness for the study population. The UK branch of the EPIC population, for which the FFQ used in the present study was derived, was deemed to be appropriate for this study because of similarities in population demographic. However, future studies should consider their study populations and adjust FFQs accordingly to capture regional and ethnic foods, as has been validated in [36–38], or add adjustment based on race and geography [39]. Furthermore, a key socioeconomic factor to consider in the interpretation of FFQ data is education status, as this has been shown to increase inaccuracy of the FFQ reporting [36, 40]. An adjusted FFQ that considers these factors could be used to create the HEI using methodology as described here, providing the use of an adequate food composition database. Alternatively, iterations of the HEI have been created and validated that better capture dietary data from other populations [16]. Similarly, the cohorts older age and majority female gender may impact how accurately FFQ captures the cohorts diet and therefore may influence the extent of variance captured by the HEI. Future studies should seek to confirm that with adjustments to the HEI, it remains the most appropriate index of choice across different populations.

A drawback of using any self-reported diet data is that individuals have a tendency to inaccurately report their consumption of food items; generally, over reporting fruits and vegetables and under reporting food items that are considered unhealthy [41, 42]. Drawing direct associations between a disease and individual dietary components derived from FFQs has been shown to be problematic [41], and points to the strength of comparative summary measures. Another consideration is that FFQ is designed to measure long-term habitual food intakes, whereas the eating behaviour immediately before microbiota sampling may diverge from typical eating behaviour. Although a lasting shift in community structure is unlikely from short-term changes [43], future studies should also explore and consider secondary adjustments to capture short-term food intakes.

The HEI as presented here had some limitations. The HEI (and MDS) were both developed in countries outside of this cohort, and an index created using UK-specific thresholds of consumption may have performed even better within this population. The HEI might also benefit from the use of a diversity measure as one of the components as has been done in adjusted HEI studies [17, 44].

The benefits of a healthy diet are well known [45], and it is important to note that a healthy diet-associated microbiota may not directly drive these outcomes or be directly influenced as a result of dietary consumption. These are also influenced by indirect effects associated with a healthy lifestyle [46]. Many of the microbial associations observed here were small. This is possibly reflective of the complex intertwining factors affecting the community composition of the microbiota. However, until the nuances of these relationships with the microbiota are fully characterised, the HEI offers an effective way of capturing wider dietary information in a single, weighted, energy-adjusted variable when other factors are of interest.

The HEI is likely inappropriate as a predictor variable of differences between microbial ecosystems as poor diets that are different but score similarly may mask trends due to specific dietary constituents. Future studies should expand on existing work to probe the effect of specific dietary elements on microbiota, but in these studies, it will be important to co-vary for overall dietary health. This study supports the use of an HEI approach in such endeavours.

## Conclusions

Of the indices studied here, associations with measures of gut microbiota composition show the Healthy Eating Index (HEI) as the most appropriate, previously established, dietary index to utilise as a covariate in microbiota studies within this population. Adjustment of thresholds or FFQ parameters could readily be applied in different demographics, but would need to be tested. As a single variable,

it is readily applied to a wide range of dietary data, has extensive resources provided by the USDA to aid its analysis and creation and is readily adjustable and interpretable. This will allow future research to adequately control for diet without saturation of models by the high dimensionality of diet data, allowing researchers to better interpret the effect of other environmental factors on the gut microbiota and potentially other human-microbe interaction models which necessitate adjustment for diet.

## Methods

### Study population

All individuals included in this study are part of the Twin-SUK research cohort, the UK's largest research database of mono- and dizygotic twins. Descriptive statistics for the data used here are provided in Table 1.

### Food frequency questionnaire (FFQ) data

Food frequency questionnaire (FFQ) data was collected following the EPIC-Norfolk guidelines [47], with only those answering for all 152 food groups considered for this analysis. As with any ongoing large-cohort study, the data was collected in batches; both undertaken on rolling bases. The first was undertaken predominantly in 2007 for 3370 individuals, and the second between 2014 and 2015 for 4116; 5047 unique individuals were used here. All analysis considered the score for the nearest time-point, excluding subjects with a greater than 5-year difference; subsequent microbiota analysis was undertaken on data matched to samples from 2070 individuals.

### HEI construction

All indices were constructed in RStudio [48] following relevant methodologies. The HEI was constructed following Guenther et al. 2013 guidelines [15]. The reported weekly frequency of each FFQ food item was converted to the unit recommended by the HEI guidelines (Additional file 1: Table S17). Divergences from the published methodology was the use of the 'Composition of foods integrated database' (CoFids) published by Public Health England [49], as a more appropriate look-up table for a UK-based cohort. CoFids was used where calories and proportional components of FFQ food items were required for calculating HEI components (Additional file 1: Table S17). Where available, volumetric conversions were calculated using specific gravities obtained from the USDA websites. Similar to subsequent studies [17, 50], food items were categorised into HEI components based on their predominant attributes (e.g. broken down into lean and solid fat fractions) after being initially classified into the USDA sub-groupings, rather than being considered within all groups (Additional file 1: Table S17.)

### MDS construction

There is some variance in methodology in assigning the MDS depending on the different weighting of evidence for factors considered to constitute a healthy diet [19, 51]; here, the MDS was constructed using the modified MDS methodology outlined by Trichopoulou et al. 2005 [19]. Estimates of daily grammes of consumption were created from residual energy-adjusted FFQ data of seven groups (Additional file 1: Table S18). Scores were assigned to each category as either 0 (no MDS) or 1 (MDS) for each category depending on whether the twin was above or below median intake of the study population. Medians were calculated using the combined scores for each FFQ 'batch'.

### HFD-index construction

Methodology from Drescher et al. 2007 [22] was used to create this index, where a healthy food value is calculated for each FFQ food item and used as a weighting for multiplication against a Simpson's index score of all consumed foods, indicating the diversity (Additional file 1: Table S19). This results in a diversity measure of diet that considers the health value of the variety of foods combined.

### Validation of indices

All statistical analysis was undertaken in Rstudio. Construct validity of indices was assessed following partial methodology from Guenther et al. 2014 [18]. First, via review of overall distributions of the total index score. Secondly, as healthy diets distinguish smokers from non-smokers, young from old and men from women, concurrent criterion validity was assessed using two sample *t* tests for the MDS and HEI, where distributions approached normalcy, and Wilcoxon. Age was calculated as age at questionnaire submission and separated into two groups; below 50s and over 50s, and those who self-reported as current smokers used to assess differences between smokers and non-smokers. 5047 individuals were used to assess age and sex differences; due to data absent due to longitudinal differences in data collection, a subset of 3226 were used to assess differences between smokers and non-smokers.

Indices were also assessed as the primary explanatory variable against health measures in nested linear regression models with age, twin zygosity, and sex as covariates against BMI on a subset of 4428 individuals missing data, again due to differences in collection method. Similarly, on a subset of 4553 individuals following the Rockwood method [24], a frailty index of the TwinsUK participants was used to indicate the health predictive capacity of each dietary indices, zygosity and sex as covariates.

### Microbiota analysis

A subset of 2070 individuals were used to assess the extent the variation within and between individuals' microbiota

could be captured by each dietary index. Collection and processing of samples for 16S rRNA gene sequencing for the TwinsUK cohort has been described previously [52]. Individuals brought samples to clinical visit or posted them in sealed ice packs to the research department where they were stored at  $-80^{\circ}\text{C}$ , until shipped frozen for analysis. DNA was extracted at Cornell University, where the V4 region of the 16S rRNA genes was amplified. A multiplexed approach was used to sequence the amplicons on the Illumina MiSeq platform. Following demultiplexing, sample read paired-ends were merged using a 200 nt minimum overlap. 16S rRNA gene sequencing data was processed and OTUs generated as described previously [53]; per sample de novo identification and removal of chimeric sequences was undertaken using USEARCH, and then de novo OTUs were picked in QIIME using SUMACLUSt at a similarity threshold of 97% [54]. The OTU representative sequences were aligned using the `parallel_align_seqs_pynast` command within QIIME, the resulting alignment was then filtered to remove variable regions using the `filter_alignment` command, and a phylogenetic tree was created using the `make_phylogeny` command. All commands were run with the default parameters in QIIME version 1.9.1.

Alpha diversity metrics of Shannon diversity, `chao1`, Simpson's diversity and observed species were also calculated in QIIME. OTUs were rarefied to 10,000 sequences per sample 50 times, and the 4 alpha diversity metrics were then calculated as the mean for each sample across the 50 rarefied tables. Mixed-effects models were constructed using the "lme4" package in R to assess the extent alpha diversity varied with dietary index; all model variables were scaled prior to input, and all reported coefficients are standardised [25]. Nested models were used to compare the effect of each dietary index. Models were adjusted for age, BMI, twin zygosity, sex and OTU count per samples, with technical covariates and FFQ questionnaire batch as random effects. As  $\chi^2$  values resulting from ANOVA of two mixed models are only appropriate for comparisons of nested models, to assess relative goodness of fit of the three dietary indices, *t* values, AICs and  $\beta$  coefficients from the mixed-effects models for each index were used to quantify the ability of a dietary index to capture each measure. To further assess the ability of dietary indices to capture variance, hierarchical models of alpha diversity were performed with BMI and a smaller subset ( $n = 2015$ ) incorporating frailty data.

Relative abundances of OTUs found in > 25% in individuals were log<sub>10</sub> transformed, and residuals were generated via regression against technical covariates of sequencing depth, sequence run, person who extracted the DNA, person who loaded the DNA and sample collection method. OTUs were collapsed to taxonomic abundances and Family and Genus levels. All OTU metrics were used



as response variables in mixed-effects models (as above) adjusted for age, twin zygosity, BMI and sex, with FFQ batch as a random effect. Nested models were compared using ANOVA, and *p* values were false discovery rate (FDR) adjusted using the *q*value package [55]. Twin pairs discordant by greater than one standard deviation and within different quartiles were identified, and OTU differences between the two were assessed using paired Wilcoxon rank-sum tests and FDR adjustment.

Unweighted UniFrac distances were calculated as  $\beta$  diversity measures using the phyloseq package in R [56]. Ordination plots were also generated using phyloseq, and the first 10 components from the PCoA (representing the first 10 axes) were extracted and used as the response variable in mixed-effects models, as in alpha diversity analysis. Finally, weighted UniFrac distances between twin pairs were used as the response variables in regression models with difference in dietary index, difference in BMI, and differences in factorial technical variables (person who extracted the DNA, person who loaded the DNA and sample collection method) as covariates. Standardised coefficients were calculated using the lm.beta package [57].

## Additional files

**Additional file 1:** Supplementary tables. (DOCX 167 kb)

**Additional file 2:** HEI creation. (ZIP 233 kb)

## Acknowledgements

The authors of this paper wish to express our appreciation to all study participants of the TwinsUK cohort for donating their samples and time. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

We thank Dr. Julia K Goodrich, Dr. Ruth E Ley and the Cornell technical team for generating the microbial data, and Serena Verdi for assistance on the frailty data.

## Funding

This study was funded by a grant from Chronic Disease Research Foundation and the Wellcome Trust (grant WT081878MA), and was supported by Clinical Research Facility at Guy's and St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust, King's College London and the University of East Anglia.

## Availability of data and materials

The European Bioinformatics Institute (EBI) accession numbers for the sequences reported in this paper is ERP015317. Individual level raw phenotype data from TwinsUK is available via application to the TwinsUK data access committee. Information on data access and how to apply is available at <http://www.twinsuk.ac.uk/data-access/submission-procedure-2/>. Supplementary material for creation of dietary indices (to be run with dietary index creation R files HEI\_creation.R, HFD\_creation.R, MDS\_creation.R for the HEI, HFD-index and MDS (Additional file 2) respectively in conjunction to FFQ derived data) has been provided as follows:

- "proximates.csv"-altered for import version of proximates tab of the McCance and Widdowson's 'composition of foods integrated dataset' on the nutrient content of the UK food supply—available in its

original form at <https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid>.

- "Food\_portions\_all.csv"-Conversions between FFQ frequency to weighted portion size, cross-referenced with relevant CoFids look up code(s) and proportion of CoFids item used to estimate for each food item.

Additionally, the R code for assessing OTU differences between twins discordant for their dietary index value is provided in discordant\_twins\_OTUs.R.

The EPIC-Norfolk Food Frequency Questionnaire and associated information and material can be found here <http://www.srl.cam.ac.uk/epic/nutmethod/FFQii.shtml>.

## Authors' contributions

RB, MJ, TP and JS prepared and performed analysis of the data used here. All authors contributed towards writing the manuscript, and read and approved the final manuscript.

## Ethics approval and consent to participate

Favourable ethical opinion was granted by the formerly known St. Thomas' Hospital Research Ethics Committee (REC). Following restructure and merging of REC, subsequent amendments were approved by the NRES Committee London-Westminster (TwinsUK, REC ref: EC04/015); use of microbiota samples was granted NRES Committee London-Westminster (The Flora Twin Study, REC ref: 12/LO/0227).

## Competing interests

TDS is co-founder of MapMySelf Ltd. and MapMyGut Ltd.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>The Department of Twin Research, Kings College London, 3-4th Floor South Wing Block D, St Thomas' Hospital, Westminster Bridge Road, London SE1 7EH, UK. <sup>2</sup>Section of Nutrition and Metabolism, International Agency for Research on Cancer (IARC-WHO), 150 Cours Albert Thomas, 69008 Lyon, France. <sup>3</sup>Norwich Medical School 2.02, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich NR4 7TJ, UK. <sup>4</sup>Clinical Age Research Unit, Kings College Hospital Foundation Trust, London, UK.

Received: 2 August 2017 Accepted: 28 March 2018

Published online: 25 April 2018

## References

1. Levy M, Kolodziejczyk AA, Thaiss CA, Elinav E. Dysbiosis and the immune system. *Nat Rev Immunol* [Internet]. 2017 [cited 2018 Jan 14];17:219–232. Available from: <http://www.nature.com/articles/nri.2017.7>.
2. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* [Internet]. 2011;9:279–90. Available from: <https://doi.org/10.1038/nrmicro2540>.
3. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. 2012;486:222–7. Available from: <https://doi.org/10.1038/nature11053>.
4. Albenberg LG, Wu GD. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. *Gastroenterology* [Internet]. 2014;146:1564–72. Available from: <https://www.sciencedirect.com/science/article/pii/S0016508514001498>.
5. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JL. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* [Internet]. 2009;1:6ra14. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2894525/>.
6. Cresci GA, Bowden E. Gut microbiome: what we do and don't know. *Nutr Clin Pract* [Internet]. 2015;30:734–46. Available from: <http://ncp.sagepub.com/content/30/6/734.abstract>.

7. Natarajan N, Pluznick JL. From microbe to man: the role of microbial short chain fatty acid metabolites in host cell biology. *Am J Physiol Physiol* [Internet]. 2014 [cited 2018 Jan 22];307:C979–C985. Available from: <http://www.physiology.org/doi/10.1152/ajpcell.00228.2014>.
8. Scheppach W. Effects of short chain fatty acids on gut morphology and function. *Gut* [Internet]. 1994 [cited 2015 Dec 17];35:S35–S38. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1378144&tool=pmcentrez&rendertype=abstract>.
9. Everard A, Lazarevic V, Gaia N, Johansson M, Ståhlman M, Backhed F, et al. Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. *ISME J* [Internet] Nature Publishing Group. 2014 [cited 2017 Jun 13];8:2116–2130. Available from: <http://www.nature.com/doi/10.1038/ismej.2014.45>.
10. Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, Keilbaugh SA, Hamady M, Chen Y, et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* [Internet] Elsevier. 2009 [cited 2016 Oct 2];137:1716–1724.e2. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0016508509014577>.
11. Cotillard A, Kennedy SP, Kong LC, Prifti E, Pons N, Le Chatelier E, et al. Dietary intervention impact on gut microbial gene richness. *Nature* [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013 [cited 2015 Mar 11];500:585–588. Available from: <https://doi.org/10.1038/nature12480>.
12. Salim SY, Kaplan GG, Madsen KL. Air pollution effects on the gut microbiota. *Gut Microbes* [Internet] Taylor & Francis. 2014;5:215–9. Available from: <https://doi.org/10.4161/gmic.27251>.
13. Rehman A, Rausch P, Wang J, Skieceviciene J, Kiudelis G, Bhagalia K, et al. Geographical patterns of the standing and active human gut microbiome in health and IBD. *Gut* [Internet]. 2016;65:238–48. Available from: <http://gut.bmj.com/content/65/2/238.abstract>.
14. Miller GE, Engen PA, Gillevet PM, Shaikh M, Sikaroodi M, Forsyth CB, et al. Lower neighborhood socioeconomic status associated with reduced diversity of the colonic microbiota in healthy adults. *Driks A, editor. PLoS One* [Internet]. Public Library of Science; 2016 [cited 2017 Nov 15];11:e0148952. Available from: <http://dx.plos.org/10.1371/journal.pone.0148952>.
15. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HAB, Kuczynski KJ, et al. Update of the healthy eating index: HEI-2010. *J Acad Nutr Diet* [Internet]. 2013 [cited 2016 Oct 5];13:569–580. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23415502>.
16. Lassale C, Gunter MJ, Romaguera D, Peelen LM, Van der Schouw YT, Beulens JWJ, et al. Diet quality scores and prediction of all-cause, cardiovascular and cancer mortality in a pan-European cohort study. *Chiu C-J, editor. PLoS One* [Internet] Public Library of Science; 2016 [cited 2016 Jul 15];11:e0159025. Available from: <http://dx.plos.org/10.1371/journal.pone.0159025>.
17. Roy R, Hebden L, Rangan A, Allman-Farinelli M. The development, application, and validation of a healthy eating index for Australian adults (HEIFA—2013). *Nutrition*. 2016;32:432–40.
18. Guenther PM, Kirkpatrick SI, Reedy J, Krebs-Smith SM, Buckman DW, Dodd KW, et al. The Healthy Eating Index-2010 is a valid and reliable measure of diet quality according to the 2010 dietary guidelines for Americans. *J Nutr American Society for Nutrition*. 2014;144:399–407.
19. Trichopoulou A, Orfanos P, Norat T, Bueno-de-Mesquita B, Ocké MC, Peeters PH, et al. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. *BMJ*. 2005;330.
20. Del Chierico F, Vernocchi P, Dallapiccola B, Putignani L. Mediterranean diet and health: food effects on gut microbiota and disease control. *Int J Mol Sci* [Internet] Multidisciplinary Digital Publishing Institute; 2014 [cited 2016 Oct 5];15:11678–11699. Available from: <http://www.mdpi.com/1422-0067/15/7/11678/>.
21. Kant AK, Block G, Schatzkin A, Ziegler RG, Nestle M. Dietary diversity in the US population, NHANES II, 1976–1980. *J Am Diet Assoc* [Internet]. 1991 [cited 2016 Sep 11];91:1526–1531. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1960344>.
22. Drescher LS, Thiele S, Mensink GBM. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J Nutr* [Internet]. American Society for Nutrition; 2007 [cited 2017 Apr 10];137:647–651. Available from: <http://jn.nutrition.org/content/137/3/647.long>.
23. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. 2012;488:178–84. Available from: <https://doi.org/10.1038/nature11319>.
24. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr* [Internet] BioMed Central; 2008 [cited 2018 Jan 5];8:24. Available from: <http://bmgeriatr.biomedcentral.com/articles/10.1186/1471-2318-8-24>.
25. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*. 2008;59:390–412.
26. Tategawkar SA, Bandinelli S, Bandeen-Roche K, Chen P, Milanecchi Y, Tanaka T, et al. A higher adherence to a Mediterranean-style diet is inversely associated with the development of frailty in community-dwelling elderly men and women. *J Nutr* [Internet]. 2012 [cited 2017 May 13];142:2161–2166. Available from: <http://jn.nutrition.org/cgi/doi/10.3945/jn.112.165498>.
27. Steyn N, Nel J, Nantel G, Kennedy G, Labadarios D. Food variety and dietary diversity scores in children: are they good indicators of dietary adequacy? *Public Health Nutr* [Internet]. 2006 [cited 2017 Jul 25];9. Available from: [http://www.journals.cambridge.org/abstract\\_S1368980006001054](http://www.journals.cambridge.org/abstract_S1368980006001054).
28. Hatloy A, Torheim LE, Oshaug A. Food variety—a good indicator of nutritional adequacy of the diet? A case study from an urban area in Mali, West Africa. *Eur. J. Clin. Nutr* [Internet]. 1998 [cited 2017 Jul 25];52:891–898. Available from: <http://www.nature.com/doi/10.1038/sj.ejcn.1600662>.
29. Jackson MA, Jeffery IB, Beaumont M, Bell JT, Clark AG, Ley RE, et al. Signatures of early frailty in the gut microbiota. *Genome Med* [Internet]. 2016 [cited 2017 Apr 13];8:8. Available from: <http://genomemedicine.com/content/8/1/8>.
30. Pallister T, Jackson MA, Martin TC, Glastonbury CA, Jennings A, Beaumont M, et al. Untangling the relationship between diet and visceral fat mass through blood metabolomics and gut microbiome profiling. *Int J Obes* [Internet]. 2017 [cited 2017 Jul 12];41:1106–1113. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28293020>.
31. Del Chierico F, Nobili V, Vernocchi P, Russo A, Stefanis C De, Gnani D, et al. Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* [Internet]. 2017 [cited 2017 Apr 13];65:451–464. Available from: <http://doi.wiley.com/10.1002/hep.28572>.
32. Tims S, Derom C, Jonkers DM, Vlietinck R, Saris WH, Kleerebezem M, et al. Microbiota conservation and BMI signatures in adult monozygotic twins. *ISME J* [Internet]. 2013 [cited 2017 Apr 13];7:707–717. Available from: <http://www.nature.com/doi/10.1038/ismej.2012.146>.
33. O'Toole PW, Jeffery IB. Gut microbiota and aging. *Science* (80- ) [Internet]. 2015 [cited 2017 May 1];350:1214–1216. Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.aac8469>.
34. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiota in new-onset Crohn's disease. *Cell Host Microbe* [Internet]. 2014 [cited 2016 Oct 8];15:382–392. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24629344>.
35. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* [Internet]. 2012 [cited 2017 Apr 13];6:320–329. Available from: <http://www.nature.com/doi/10.1038/ismej.2011.109>.
36. Mayer-Davis EJ, Vitolins MZ, Carmichael SL, Hemphill S, Tsaroucha G, Rushing J, et al. Validity and reproducibility of a food frequency interview in a multi-cultural epidemiologic study. *Ann Epidemiol* [Internet]. 1999 [cited 2018 Jan 14];9:314–324. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1047279798000702>.
37. Neelakantan N, Whitton C, Seah S, Koh H, Rebello S, Lim J, et al. Development of a semi-quantitative food frequency questionnaire to assess the dietary intake of a multi-ethnic urban Asian population. *Nutrients* [Internet]. 2016 [cited 2018 Jan 14];8:528. Available from: <http://www.mdpi.com/2072-6643/8/9/528>.
38. Ishihara J, Iwasaki M, Kunieda CM, Hamada GS, Tsugane S. Food frequency questionnaire is a valid tool in the nutritional assessment of Brazilian women of diverse ethnicity. *Asia Pac J Clin Nutr* [Internet] HEC Press; 2009 [cited 2018 Jan 17];18:76–80. Available from: <http://www.airitilibrary.com/Publication/alDetailedMesh?docid=09647058-200903-201306130010-201306130010-76-80>.
39. Signorello LB, Munro HM, Buchowski MS, Schlundt DG, Cohen SS, Hargreaves MK, et al. Estimating nutrient intake from a food frequency questionnaire: incorporating the elements of race and geographic region. *Am J Epidemiol* [Internet] Oxford University Press; 2009 [cited 2018 Jan 14];170:104–111. Available from: <https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwp098>.
40. Hébert JR, Peterson KE, Hurley TG, Stoddard AM, Cohen N, Field AE, et al. The effect of social desirability trait on self-reported dietary measures

- among multi-ethnic female health center employees. *Ann. Epidemiol.* [Internet]. 2001 [cited 2018 Jan 14];11:417–427. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1047279701002125>.
41. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, Bingham S, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based observing protein and energy nutrition (OPEN) study. *Int J Epidemiol* [Internet] Oxford University Press; 2003 [cited 2016 Oct 2];32:1054–1062. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14681273>.
  42. Shu XO, Yang G, Jin F, Liu D, Kushi L, Wen W, et al. Validity and reproducibility of the food frequency questionnaire used in the Shanghai Women's Health Study. *Eur J Clin Nutr* [Internet]. 2004 [cited 2016 Oct 11];58:17–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14679362>.
  43. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial Enterotypes. *Science* [Internet]. 2011;334:105–8. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3368382/>.
  44. Hann CS, Rock CL, King I, Drewnowski A. Validation of the healthy eating index with use of plasma biomarkers in a clinical sample of women. *Am J Clin Nutr* [Internet]. 2001 [cited 2016 May 10];74:479–486. Available from: <http://ajcn.nutrition.org/content/74/4/479.full>.
  45. Kaput J, Rodriguez RL. Nutritional genomics: the next frontier in the postgenomic era. *Physiol Genomics* [Internet]. 2004 [cited 2017 Mar 19];16:166–177. Available from: <http://physiolgenomics.physiology.org/lookup/doi/10.1152/physiolgenomics.00107.2003>.
  46. Hanage WP. Microbiology: microbiome science needs a healthy dose of scepticism. *Nature* [Internet]. 2014 [cited 2017 Mar 19];512:247–248. Available from: <http://www.nature.com/doi/10.1038/512247a>.
  47. EPIC-Norfolk nutritional methods: food frequency questionnaire [Internet]. [cited 2017 May 13]. Available from: <http://www.srl.cam.ac.uk/epic/nutmethod/FFQ.shtml>.
  48. RStudio Team. RStudio: integrated development for R. [internet]. Boston, MA: RStudio, Inc.; 2015. Available from: <http://www.rstudio.com/>.
  49. Finglas P, Roe M, Pinchen H, Berry R, Church S, Dodhia S, et al. Composition of foods integrated dataset user guide 2 about Public Health England composition of foods integrated dataset user guide. 2015 [cited 2016 Oct 3]; Available from: <https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid>.
  50. McNaughton SA, Ball K, Crawford D, Mishra GD. An index of diet and eating patterns is a valid measure of diet quality in an Australian population. *J Nutr* [Internet]. 2008 [cited 2016 May 10];138:86–93. Available from: <http://jn.nutrition.org/content/138/1/86.short>.
  51. Liese AD, Krebs-Smith SM, Subar AF, George SM, Harmon BE, Neuhouser ML, et al. The dietary patterns methods project: synthesis of findings across cohorts and relevance to dietary guidance. *J Nutr* [Internet] American Society for Nutrition; 2015 [cited 2016 Oct 10];145:393–402. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25733454>.
  52. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekman R, et al. Human genetics shape the gut microbiome. *Cell* [Internet]. 2014 [cited 2017 May 1];159:789–799. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0092867414012410>.
  53. Jackson MA, Bell JT, Spector TD, Steves CJ. A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units. *PeerJ* [Internet] PeerJ, Inc; 2016 [cited 2017 May 1];4:e2341. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27635321>.
  54. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* [Internet] Nature Publishing Group; 2010 [cited 2016 Oct 11];7:335–336. Available from: <http://www.nature.com/doi/10.1038/nmeth.f.303>.
  55. Storey J, Bass A, Dabney A, Robinson D. Qvalue: Q-value estimation for false discovery rate control. 2015.
  56. McMurdie PJ, Holmes S, Kindt R, Legendre P, O'Hara R. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. Watson M, editor. *PLoS One* [Internet]. Public Library of Science; 2013 [cited 2017 Apr 10];8:e61217. Available from: <http://dx.plos.org/10.1371/journal.pone.0061217>.
  57. Stefan Behrendt. Im.beta: Add standardized regression coefficients to Im-Objects. 2015 [cited 2018 Jan 23]; Available from: <https://cran.r-project.org/web/packages/Im.beta/Im.beta.pdf>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

