



The technique ‘joint and individual variance explained’ highlights persistent aspects of the diet using longitudinal food frequency data

M. Beatrix Jones^{1*}, Amaan Merchant¹, Larisa Morales-Soto², John M. D. Thompson³ and Clare R. Wall⁴

¹Department of Statistics, Faculty of Science, University of Auckland, Auckland 1142, New Zealand

²Department of Human Genetics, McGill University, Montreal, Canada

³Department of Paediatrics, Child and Youth Health, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

⁴Department of Nutrition and Dietetics, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

(Submitted 16 August 2021 – Final revision received 10 November 2021 – Accepted 7 December 2021 – First published online 17 December 2021)

Abstract

Dietary pattern analysis is typically based on dimension reduction and summarises the diet with a small number of scores. We assess ‘joint and individual variance explained’ (JIVE) as a method for extracting dietary patterns from longitudinal data that highlights elements of the diet that are associated over time. The Auckland Birthweight Collaborative Study, in which participants completed an FFQ at ages 3–5 (*n* 549), 7 (*n* 591) and 11 (*n* 617), is used as an example. Data from each time point are projected onto the directions of shared variability produced by JIVE to yield dietary patterns and scores. We assess the ability of the scores to predict future BMI and blood pressure measurements of the participants and make a comparison with principal component analysis (PCA) performed separately at each time point. The diet could be summarised with three JIVE patterns. The patterns were interpretable, with the same interpretation across age groups: a vegetable and whole grain pattern, a sweets and meats pattern and a cereal *v.* sweet drinks pattern. The first two PCA-derived patterns were similar across age groups and similar to the first two JIVE patterns. The interpretation of the third PCA pattern changed across age groups. Scores produced by the two techniques were similarly effective in predicting future BMI and blood pressure. We conclude that when data from the same participants at multiple ages are available, JIVE provides an advantage over PCA by extracting patterns with a common interpretation across age groups.

Key words: Dietary patterns: Principal components analysis: Joint and individual variance explained: Longitudinal data: FFQ: Dimension reduction

A number of longitudinal studies involving the identification and extraction of dietary patterns have taken place over the last 15 years, with the popularity of these studies increasing following a 2005 paper by Mikkilä *et al.*⁽¹⁾ Dietary patterns characterise food consumption trends across a population, and their associated scores provide a summary of an individual's diet. Patterns and scores can be obtained by various methods, most notably reduced rank regression^(2–11), or some form of principal components analysis (PCA) or factor analysis^(1,12–23). Other methods such as cluster analysis^(24,25), latent class analysis⁽²⁶⁾ and scoring on *a priori* patterns^(27,28) have also been used.

In this literature, populations considered included adults^(2,4–6,13,17,18,24,26,27) and children and adolescents^(3,7,11,15,16,19,21,22,25,29,30). A mix of FFQ^(3,5,7,8,12,14,17,20,22,24,26), recalls⁽¹³⁾ and food diaries^(2,4,6,9–11,18,27) was used. Some studies extracted different dietary patterns at each time point^(2,7,8,12,14–17,20–22,24,25,27), while

others used one time point to calculate the food-specific loadings for each dietary pattern and then applied these across all time points to calculate scores^(3–6,9–11,13,18,19). One study⁽²⁶⁾ combined the data from different time points into one pooled set before extracting dietary patterns. Many of these studies examine the predictability or ‘tracking’ of dietary pattern scores and individual foods, either through correlations across pairs of time periods^(1,3,8,13,14,21,22), using generalised estimating equations to get a measure of consistency across all time periods^(5,10,19), or examining whether individuals remain in the same cluster or quartile of consumption^(7,19,24,25,29). While methodological differences make comparison across studies difficult, Mikkilä *et al.*⁽¹⁾ noted stronger tracking for individuals that were greater than 15 years of age at the beginning of their study. The Avon Longitudinal Study⁽²³⁾ found changes in dietary patterns between 3- and 4-year-olds, and between 7- and 9-year-olds.

Abbreviations: JIVE, joint and individual variance explained; PCA, principal component analysis; FFQ, food frequency questionnaire; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index.

* **Corresponding author:** Dr M. B. Jones, email beatrix.jones@auckland.ac.nz

In this present study, we evaluated the utility of joint and individual variance explained (JIVE)⁽³¹⁾ for deriving longitudinal dietary patterns. JIVE is an extension of PCA for multiple datasets that extracts dimensions maximising shared variability across the datasets. JIVE shares the advantages of approaches that derive dietary patterns separately at each age point, and approaches that extract patterns from a single dataset and score the other time points for these patterns. As in methods that extract patterns separately for each time point, changes in the questionnaire across time points are easily accommodated, as are changes in the number of subjects due to incomplete follow-up. As in methods where the patterns are derived from a single age, the scores at each age have a common interpretation. A novel feature of JIVE patterns is that by prioritising shared variance, the patterns produced preferentially include aspects of the diet that are predictable across time periods. This includes both foods with high tracking coefficients and foods that are predictable from consumption of different foods in other time periods. This ability is particularly relevant for studies of childhood where children's changing capabilities (reduced choking risk for older children, use of knife and fork, increasing individual control of food choices) naturally change the range of foods consumed.

We aimed to illustrate the JIVE methodology using the FFQ responses from children at ages 3-5 and 7 from a New Zealand birth cohort study⁽²²⁾, with the addition of data at age 11. We also examined the association of JIVE scores from younger ages with BMI and blood pressure at future time points, to assess whether aspects of the diet captured had any association with health outcomes. We envision JIVE will be used in situations where researchers would like an exploratory analysis of the diet, rather than one targeted to particular nutrients. PCA is also appropriate to these circumstances, and widely used, so it is computed for comparison.

Methods

PCA can be thought of as a low-dimensional reconstruction of the original n by p data matrix \mathbf{D} using a small number k of orthogonal directions \mathbf{V} and scores \mathbf{S} .

$$\mathbf{D} = \mathbf{V}\mathbf{S} + \boldsymbol{\varepsilon}$$

For a specified k , \mathbf{V} and \mathbf{S} minimise the sum of squared errors, $\sum_{ij} \varepsilon_{ij}^2$.

JIVE⁽³¹⁾ expands this representation to multiple datasets with shared variability. In our case, these are datasets at different time points t ranging from 1 to T . The shared variability is represented by k dimensional time point-specific direction sets $\mathbf{V}_1 \dots \mathbf{V}_T$, with common scores \mathbf{S} . Time point-specific variation for time point t is represented by directions \mathbf{U}_t (constrained to have orthogonal columns and to be orthogonal to \mathbf{V}_t) and time point-specific scores \mathbf{R}_t , where the dimension of \mathbf{U}_t , l_t , may vary with time point. Thus, the dataset at time point t can be represented as

$$\mathbf{D}_t = \mathbf{V}_t\mathbf{S} + \mathbf{U}_t\mathbf{R}_t + \boldsymbol{\varepsilon}$$

The fitting procedure again minimises $\sum_{ij} \varepsilon_{ij}^2$, subject to pre-specified k , $l_1 \dots l_T$.

While the joint score vector \mathbf{s}_t for individual I represents a summary of that individual in the shared variation space and is useful for (e.g.) identifying groups of similar individuals, it represents a compromise score over all time points and cannot be computed without complete data over all time points. We would like to be able to compute a score for all individuals at each time point, for example, an individual with data at time 1 only. Thus, we worked with the projection of the data from time point t , \mathbf{D}_t , onto the shared variability directions \mathbf{V}_t : $\mathbf{S}_t = \mathbf{D}_t\mathbf{V}_t$. As in PCA, the amount of variability in \mathbf{D}_t represented by this projection can be calculated, and the correlation of the scores \mathbf{S}_t with the original variables can be used to interpret the directions, analogous to PCA loadings. Thus, although only individuals with food data at all time points were used to determine the directions, scores on those directions were computed for every individual with food data at time t , and all of these individuals were considered when computing the proportion of variance explained and interpreting the derived patterns. The correlations across time points for the scores from projecting onto the i -th column of each \mathbf{V}_t represent the strength of the inter-time point associations being represented with the i -th set of directions. In contrast with PCA and the global summary scores \mathbf{S} , the set of scores \mathbf{S}_t are not perfectly uncorrelated.

To choose k and $l_1 \dots l_T$, we followed the permutation procedure outlined in Lock *et al.*⁽³¹⁾. However, when interpreting the patterns, we selected the number of patterns to consider using the proportion of variance explained for the full complement of individuals observed at each time point.

JIVE was developed for distinct data types, rather than time points, that is, the variables in set i were completely different from those in set j . This means any change in the diet survey across time points is not a barrier to using JIVE, although in the longitudinal setting we expect many shared variables. Visualisations can highlight the changing role of particular variables over time.

We computed the JIVE directions and examined their interpretation for an FFQ completed as part of the Auckland Birthweight Collaborative Study⁽²²⁾ at ages 3-5, 7 and 11. Study participants were born between October 1995 and November 1997; approximately 50% of participants were born small for gestational age (\leq 10th percentile for sex and gestation) and 50% appropriate for gestational age. Follow-up at ages 7 and 11 was restricted to children of mothers of European descent, due to inadequate sample sizes for other groups, so only the European descent group was considered here. Ethical approval for each phase of this study was obtained from the Northern Regional Ethics Committee. Parents of participating children gave written informed consent, and 11-year-old children gave assent at the assessment. There were 549 participants at age 3-5, 591 at age 7 and 617 at age 11. The number of FFQ items was, respectively, 88, 97 and 109, with eighty-two items consistently in the FFQ. Table 1 shows the number of individuals with FFQ data at each age and the overlap between participants at different ages. The questionnaire items for each age are available in online Supplementary Table S1.

FFQ items were recorded on a 0-7 scale ranging from 'never' to '2 or more times per day'. A version of the FFQ for infants (6-24 months) was validated against a 4-day food record

Table 1. The number of individuals completing the FFQ at each age (diagonal entries), and shared individuals across surveys

	Age 3-5	Age 7	Age 11
Shared with age 3-5	549		
Shared with age 7	478	591	
Shared with age 11	480	547	617
Shared across all ages		466	

and also showed good short-term repeatability⁽³²⁾. The questionnaire was adapted to include foods eaten by older children, with reference to the New Zealand 2002 National Children's Nutrition Survey⁽³³⁾. Response variables were standardised within each age group prior to analysis. We compared with conventional PCA scores computed at each of these ages; note that the current analysis differs from previous analyses of the Auckland Birthweight Collaborative data⁽²²⁾ in that no rotation was performed, and a wider selection of survey items was included. We required JIVE and PCA to use the same number of components for comparability. We then compared the interpretation of the PCA and projected JIVE scores, and their ability to predict health measurements at future time points. The health outcomes assessed were BMI Z-score, based on age- and sex-specific standards⁽³⁴⁾ and systolic blood pressure (SBP) and diastolic blood pressure (DBP). The current value of the relevant health metric was used as a covariate (e.g. age 7 BMI Z-score was predicted by age 3-5 BMI Z-score plus dietary pattern scores). For both SBP and DBP, the Z-score for height at the earlier age, and sex, were also used as covariates, because they have been shown to influence blood pressure in children⁽³⁵⁾. The utility of the dietary patterns was assessed by a likelihood ratio test comparing the model with the covariates only, and the covariates plus dietary pattern scores.

Models where the diet significantly impacted the response variable were then assessed in more detail, for both the JIVE and PCA patterns. The significance of the individual pattern scores in a linear regression model was examined. All analyses are performed in R version 4.0.2⁽³⁶⁾ and associated packages⁽³⁷⁻⁴⁰⁾; *P* values < 0.05 are considered significant throughout.

Results

The proportion of variance explained by each JIVE or PCA pattern is shown in Fig. 1. PCA, by construction, maximises the variance explained, but projection onto the JIVE directions explains a similar amount of variability. An 'elbow' criteria suggest 1-3 patterns across the different ages and methods; three patterns were selected to maximise the amount of variability represented.

The correlations between the original FFQ items and the scores on the age-specific directions for each pattern, for both JIVE and PCA, are shown in Fig. 2. A food is shown if, for either technique, the absolute correlation between that food and the age-specific score is larger than 0.35 for at least one age group. Where there was a change in the questionnaire, a food appears only in the relevant shades. For instance, only the age 11 questionnaire had an item about avocado consumption.

The first two derived patterns are similar across JIVE and PCA. The first pattern, 'vegetables and whole grain', has positive correlations, for all ages, with vegetables, whole grains, fish and nuts, and a negative correlation with white bread. The second pattern, 'sweets and meats', has positive correlations with a variety of sweet beverages and snack foods, as well as chops/roast, hamburgers, bacon/ham and mixed dishes with meat.

The third pattern is different between the two techniques. For JIVE, the pattern has two poles, representing a trade-off between cereal, milk on cereal, margarine and selected vegetables at one end and sweet drinks and lollies at the other. The third PCA pattern does not have a common interpretation across the ages. Cereal, milk on cereal and negative associations with berries and soft drinks were most highly associated with the third principal component score for 3-5-year-olds; margarine and a variety of vegetables had the highest association for 7-year-olds; and carrots, apples and a negative association with fish canned in oil were the strongest associations for 11-year-olds.

The discordant PCA patterns are reflected in the correlations of the pattern scores across ages, as shown in Fig. 3. The scores for the third PCA pattern are not highly associated across ages; in fact, the third score is negatively correlated between age 3-5 and age 11. In contrast, the JIVE scores have high association across all ages, by construction. The trade-off for this feature is the modest correlation between the three JIVE scores within each age group. The within-age score correlations are, by construction, zero for PCA.

The specific foods associated with a JIVE pattern can change across the age groups. For example, the third pattern prominently includes sweet drinks. The most important sweet drink for 3-5- and 7-year-olds in this pattern is juice, but this switches to soft drink for 11-year-olds. The shared variability represented by this component suggests the consumption of soft drink at age 11 is to some extent predictable by juice consumption at earlier ages. (Pattern 2, which loads heavily on soft drink for all ages, indicates that soft drink and cordial consumption at earlier ages were also associated with soft drink consumption at later ages.)

The two types of scores have almost identical ability to predict future health metrics, as measured by R^2 (shown in Table 2). The likelihood ratio test shows that the dietary pattern scores are useful predictors for BMI Z-score (regardless of the age pair considered), for the prediction of age 11 blood pressure with age 3-5 scores and for the prediction of age 7 SBP.

The models where significant effects for dietary patterns were found are given in more detail in Table 3. The food scores are standardised so that the fitted coefficients represent the estimated effect of a 1 sd change. For the models with a continuous response, a higher score on JIVE pattern 1 (vegetables and whole grains) was significantly associated with a better (lower) measurement for BMI in all cases, and blood pressure at age 11 predicted with age 3-5 data. The results for PCA pattern 1 were similar, though did not reach statistical significance for BMI age 7, BMI age 11 predicted with age 7 data and diastolic blood pressure age 11 predicted with age 3-5 data. JIVE pattern 2 (sweets and meats) had a detrimental effect on the response that is significant for the BMI models and SBP. Again, the second PCA score had a similar pattern, although significant for age 11 diastolic blood pressure and not significant for age 11 SBP.



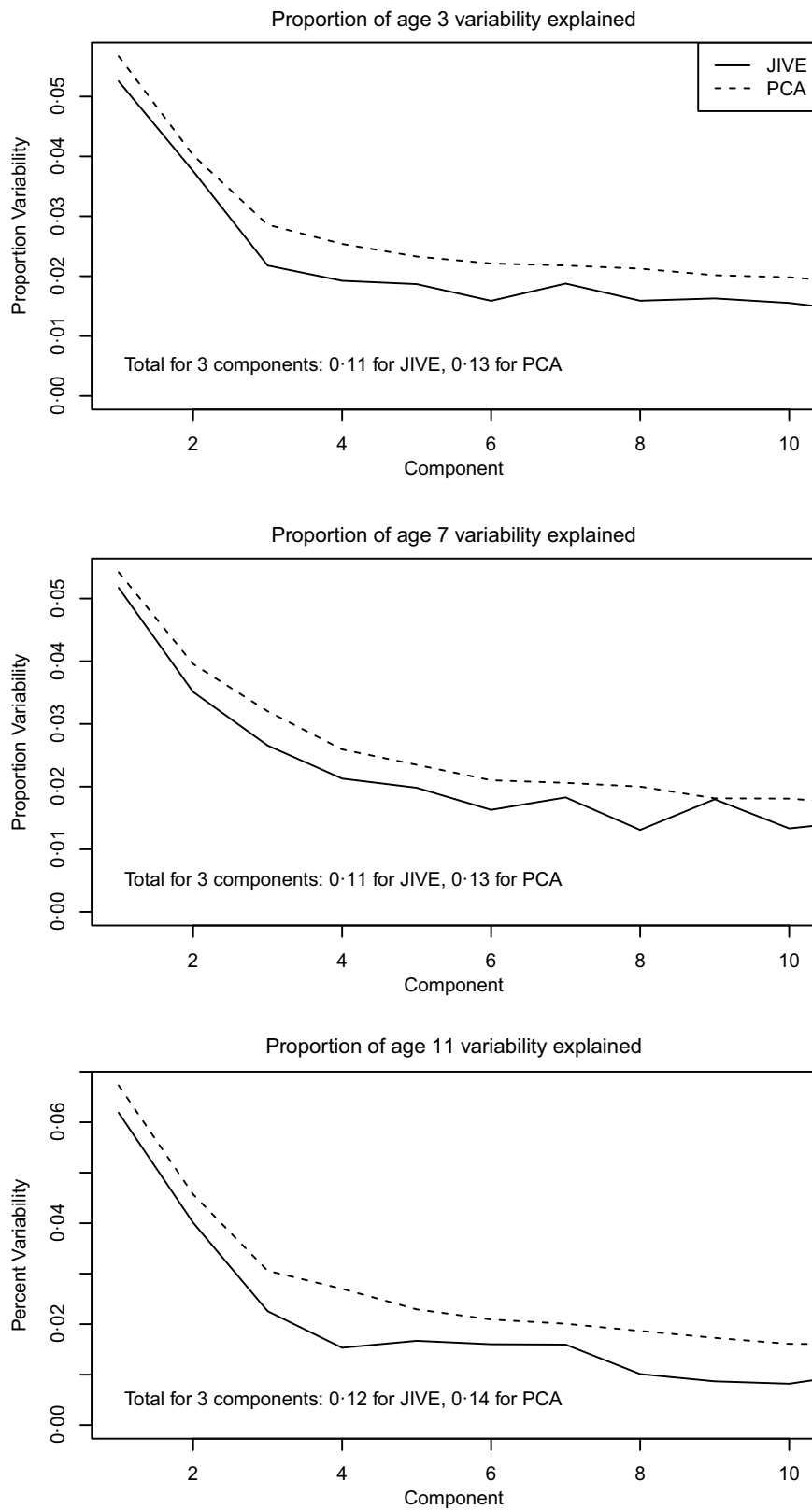


Fig. 1. Proportion of total variance explained by the scores on the JIVE directions and PCA scores. JIVE, joint and individual variance explained; PCA principal component analysis.

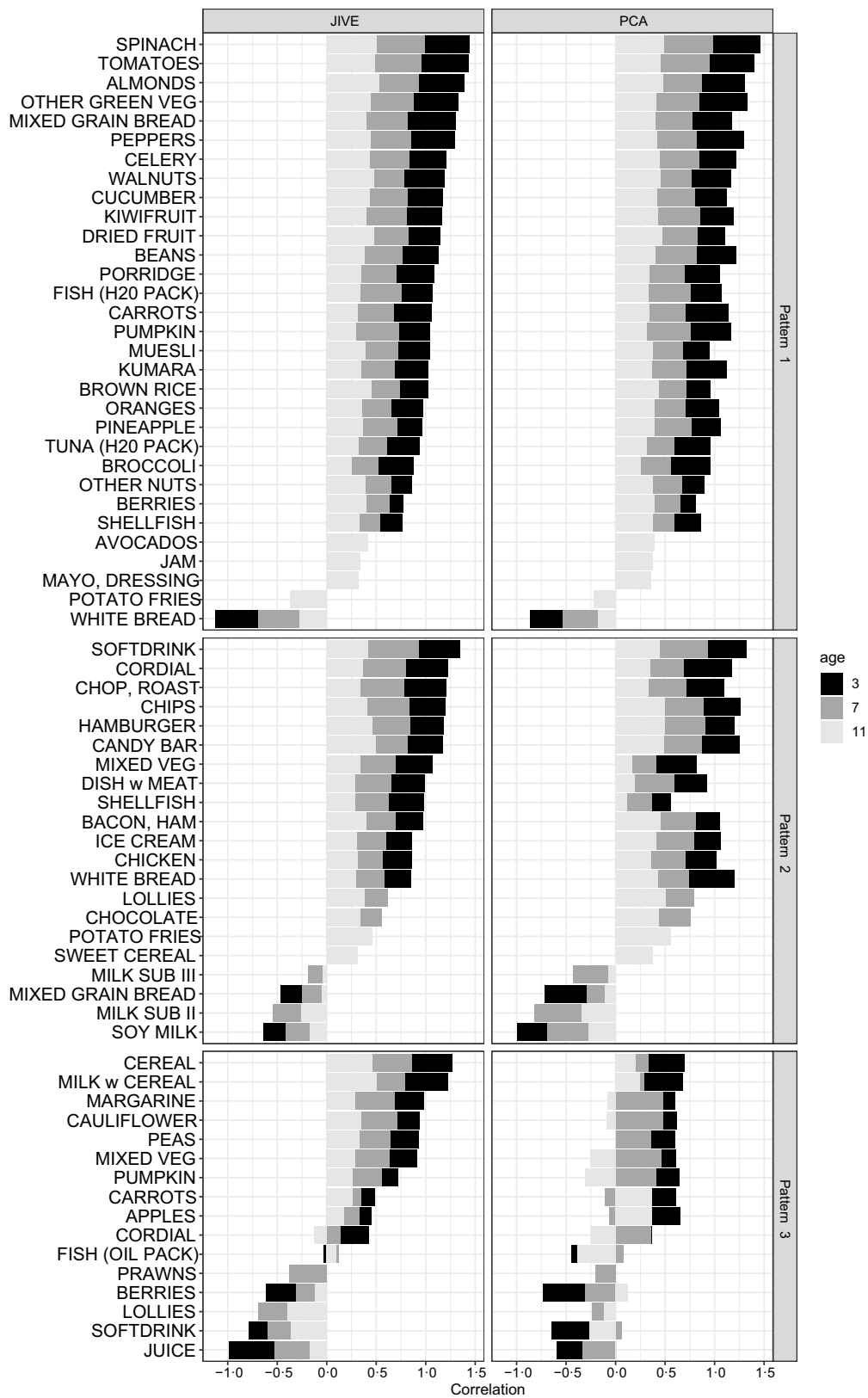


Fig. 2. Correlations between pattern scores and original variables. Correlations for different ages are stacked. Foods with an absolute correlation of > 0.35 for either JIVE or PCA, at any age, are shown. When not all shades are shown for a particular item, it indicates that the item was only assessed as a subset of the ages. JIVE, joint and individual variance explained; PCA principal component analysis.

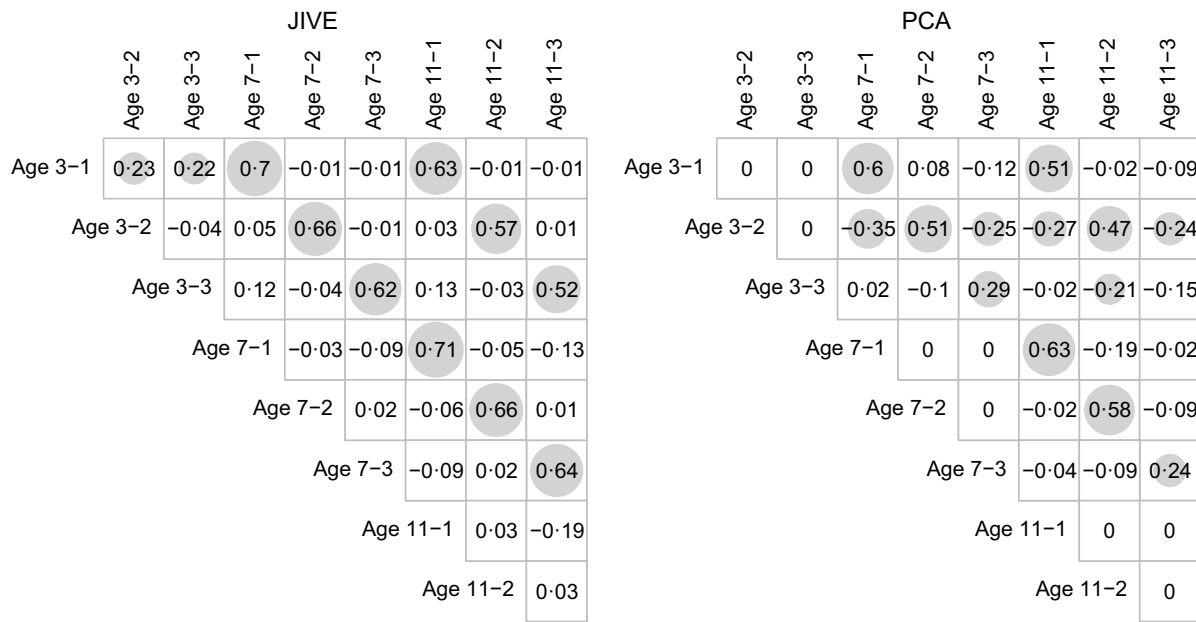


Fig. 3. Correlations between the three scores produced at each age, for JIVE and PCA. Grey circles mark correlations with absolute value larger than 0.2. JIVE, joint and individual variance explained; PCA principal component analysis.

JIVE pattern 3 (positively associated with cereal, negatively with sweet drinks) is not significant in any of the linear models. Higher scores for PCA pattern 3 for age 3-5 (positively associated with cereal, negatively with berries and soft drinks) were associated with significantly lower BMI at both ages 7 and 11.

The largest coefficient when predicting BMI Z-score is -0.17 , for JIVE pattern 1 in the model predicting age 11 BMI from age 3-5 data. In other words, scoring 1 sd higher on the 'healthy' food pattern at age 3-5 corresponds, on average, to a 0.17 lower BMI Z-score at age 11.

Discussion

The JIVE projections resulted in scores that are more interpretable across age groups than principal component scores. For instance, if we compare a model using age 3-5 predictors to one using age 7 predictions, a statement like 'pattern 3 has a positive coefficient in both models' is meaningful, because pattern 3 represents something similar in both age groups. This is not necessarily true for principal component scores derived separately for each dataset; for our data, the first two component scores had a similar interpretation across ages, but the third principal component represented disparate sets of food groups for the three ages. The third JIVE component also highlighted cereal and associated milk consumption as a food that is predictable across ages, although it represents a smaller proportion of total variability at ages 7 and 11.

Varimax rotation is commonly used in association with PCA to improve interpretability within age group; it could either improve or worsen association across age groups, depending on the dataset. A previous analysis of the Auckland Birthweight Collaborative data for ages 3-5 and 7⁽²²⁾ used PCA with varimax rotation; the patterns derived show differences from the

unrotated PCA patterns derived here. In particular, rotation grouped cereal and associated milk consumption with vegetables in what the previous authors call the 'healthy' pattern. Some meat items were grouped with a 'traditional' pattern, leaving the pattern they call 'junk' more clearly associated with sweets and snack food than the 'sweets and meats' pattern we have described with the unrotated analysis. The interpretation of the rotated PCA patterns was more similar across ages, but scores across ages were overall less correlated than those observed in the unrotated patterns (correlation between age 3-5 and age 7 scores was 0.36, 0.39 and 0.32 for the rotated patterns, and 0.60, 0.51 and 0.29 for the unrotated patterns).

A potential limitation of JIVE is that the scores for the distinct patterns are not constrained to be perfectly uncorrelated. In practice, the highest correlation observed was 0.23, between JIVE patterns 1 and 2 for the age 3-5 data. This modest correlation did not create problems with collinearity in the prediction models, and both patterns were found to be significantly associated with BMI.

We have fit relatively simple models to show that the use of JIVE scores rather than PCA scores does not obscure associations between the diet and outcome variables. However, either type of score, used in isolation, is susceptible to representing the effects of potentially confounded variables such as activity level.

A further limitation of our study is that we did not compare to the common dimension reduction technique reduced rank regression. Reduced rank regression may have better concordance across ages because each pattern extraction is supervised by common variables, typically densities of various nutrients and macronutrients. Common supervisory variables include dietary energy density, dietary fibre density, saturated fat intake, the proportion of total energy from fat and glycaemic index^(4,5,7-12). This type of approach is not relevant for the Auckland Birthweight Collaborative Study data, as the FFQ is

Table 2. Summary of linear regression models predicting health outcomes with food patterns at earlier ages

Outcome	Score type	Age 3-5 predictors			Age 7 predictors		
		Rsq	P*	n	Rsq	P	n
Age 7 BMI†	JIVE	0.63	0.015	478			
	PCA	0.63	0.001	478			
Age 11 BMI†	JIVE	0.46	0.000	480	0.73	0.009	547
	PCA	0.47	0.000	480	0.73	0.023	547
Age 7 DBP‡	JIVE	0.159	0.823	445			
	PCA	0.160	0.729	445			
Age 11 DBP‡	JIVE	0.100	0.048	437	0.170	0.214	529
	PCA	0.099	0.056	437	0.165	0.707	529
Age 7 SBP‡	JIVE	0.317	0.011	443			
	PCA	0.314	0.027	443			
Age 11 SBP‡	JIVE	0.173	0.011	435	0.241	0.082	524
	PCA	0.173	0.011	435	0.241	0.086	524

BMI, body mass index; JIVE, joint and individual variance explained; PCA, principal component analysis; DBP, diastolic blood pressure; SBP, systolic blood pressure.

* The P value given is for the likelihood ratio test comparing the models with and without the food scores. Values <0.05 are considered significant.

† BMI models include BMI at the earlier age, and three food scores computed using either JIVE or PCA.

‡ Blood pressure models include sex, the same blood pressure measurement and height at the earlier age and three food scores.

Table 3. Model coefficients for linear models where the food scores (V1–3) had a significant impact for either JIVE or PCA food scores

Response	Predictors	JIVE model			PCA model		
		Coefficient*	SE	P	Coefficient	SE	P
Age 7 BMI-Z	Intercept	0.19	0.03	0.000	0.19	0.03	0.000
	Age 3-5 V1	-0.08	0.03	0.006	-0.05	0.03	0.069
	Age 3-5 V2	0.06	0.03	0.029	0.07	0.03	0.023
	Age 3-5 V3	-0.03	0.03	0.295	-0.08	0.03	0.004
	Age 3-5 BMI-Z	1.05	0.04	0.000	1.05	0.04	0.000
Age 11 BMI-Z	Intercept	0.16	0.04	0.000	0.16	0.04	0.000
	Age 3-5 V1	-0.17	0.04	0.000	-0.11	0.04	0.003
	Age 3-5 V2	0.11	0.04	0.007	0.11	0.04	0.006
	Age 3-5 V3	-0.07	0.04	0.102	-0.15	0.04	0.000
	Age 3-5 BMI-Z	1.01	0.05	0.000	1.01	0.05	0.000
Age 11 BMI-Z	Intercept	0.01	0.03	0.838	0.01	0.03	0.852
	Age 7 V1	-0.06	0.03	0.019	-0.05	0.03	0.064
	Age 7 V2	0.06	0.03	0.013	0.06	0.03	0.035
	Age 7 V3	0.01	0.03	0.664	0.03	0.03	0.184
	Age 7 BMI-Z	0.96	0.03	0.000	0.96	0.03	0.000
Age 11 DBP mmHg	(Intercept)	44.22	2.60	0.000	68.56	4.88	0.000
	Age 3-5 V1	-0.77	0.31	0.013	-0.84	0.39	0.034
	Age 3-5 V2	0.52	0.30	0.081	0.88	0.41	0.031
	Age 3-5 V3	-0.36	0.30	0.236	-0.59	0.39	0.132
	Age 3-5 DBP	0.26	0.05	0.000	0.35	0.05	0.000
	Male sex	1.62	0.58	0.005	1.20	0.79	0.128
	Age 3-5 Height-Z	0.31	0.28	0.261	1.39	0.38	0.000
	(Intercept)	59.72	3.71	0.000	59.66	3.72	0.000
Age 7 SBP mmHg	Age 3-5 V1	-0.56	0.31	0.074	-0.20	0.30	0.504
	Age 3-5 V2	0.92	0.30	0.002	0.86	0.30	0.005
	Age 3-5 V3	-0.09	0.31	0.767	-0.25	0.29	0.395
	Age 3-5 SBP	0.41	0.04	0.000	0.42	0.04	0.000
	Male sex	1.79	0.60	0.003	1.77	0.60	0.003
	Age 3-5 Height-Z	1.48	0.28	0.000	1.46	0.28	0.000
	(Intercept)	68.29	4.88	0.000	44.16	2.60	0.000
	Age 3-5 V1	-1.26	0.42	0.003	-0.53	0.29	0.072
Age 11 SBP mmHg	Age 3-5 V2	0.81	0.40	0.044	0.48	0.3	0.106
	Age 3-5 V3	-0.51	0.41	0.215	-0.40	0.29	0.169
	Age 3-5 SBP	0.35	0.05	0.000	0.26	0.05	0.000
	Male sex	1.25	0.79	0.114	1.57	0.58	0.007
	Age 3-5 Height-Z	1.41	0.38	0.000	0.30	0.28	0.28

JIVE, joint and individual variance explained; PCA, principal component analysis; BMI-Z, Z-score for body mass index; DBP, diastolic blood pressure; SBP, systolic blood pressure.

* Food scores were standardised so that the coefficient represents the effect of a 1 SD change; responses and other covariates are on their original scales.



not validated for nutrient intake. Beyond this practical issue, selecting supervisory variables imposes limitations on what type of common structures can be found. By focussing on shared variability, JIVE-derived patterns characterise aspects of the diet that persist across ages, whatever those aspects may be.

Finally, the JIVE directions were determined using only individuals who completed FFQ at all three time points (although scores were computed for all individuals). The recently developed technique generalised integrative PCA⁽⁴¹⁾ theoretically would allow the selected directions to be influenced by individuals missing FFQ data at one or more time points; however, currently it lacks a robust software implementation.

Despite these limitations, we conclude that JIVE scores offer important practical advantages, and the possibility of discovering novel associations between the diets at different ages. JIVE score computation easily accommodates changes in the questionnaire used, rather than requiring an identical set of variables for each age.

We also believe JIVE could be useful in the validation of FFQ. In this process, a subset of individuals completes both an FFQ and food diary at the same time point, and the concordance between the two response sets is summarised. Applying JIVE, with the two instruments analogous to two 'time points,' would help summarise what broad diet patterns are captured by both instruments, and which are captured by only one. JIVE is designed to represent linear relationships, so in any situation relating FFQ to food diaries, one might consider transforming the food diary quantities so that there is a linear correspondence with the FFQ scale. For instance, for our FFQ, the scale category midpoints correspond to 0, 0.5, 2, 4, 12, 22, 28 and 56 servings/month; taking the square root of servings gives a near-linear correspondence with the 0–7 scale. Visualisations such as Fig. 2, where most food groups appear in all three assessments, would require that food diary items are grouped and named so that there is reasonable correspondence across instruments. It may also be tempting to apply JIVE opportunistically if data from different methodologies (e.g. food diary and FFQ) were available at different time points of a longitudinal study; however, interpretation would be difficult without a detailed understanding of how instrument differences contributed to any lack of concordance.

Finally, we note that the JIVE scores have the potential to represent age-related changes in how particular habits or preferences express themselves. For instance, 11-year-olds had a higher weighting for soft drinks in the 'sweet drinks' aspect of JIVE pattern 3; younger children had a higher weighting for juice. Just as reduced rank regression attempts to find dietary patterns related to a specific outcome using a biomarker to supervise pattern discovery, our use of JIVE could be seen as using the FFQ data from ages 7 and 11 to supervise the discovery of patterns in the age 3–5 data. Thus, JIVE can be used to identify dietary habits that are established at a young age. In cohorts with historic diet information, JIVE scores could also be used to characterise long-term habits and examine their association with the emergence of chronic disease. We believe this knowledge will assist in the design of early interventions to establish healthy eating behaviours.

Acknowledgements

A. M. and L. M. were supported by summer studentships from the University of Auckland Faculty of Science.

M. B. J. designed research; J. T. and C. W. provided essential data; M. B. J., A. M. and L. M. performed statistical analysis; all authors interpreted the analysis; M. B. J. and A. M. wrote the first draft of the paper; M. B. J. had primary responsibility for final content. All authors read, critically appraised and edited the paper, and approved the final manuscript.

There are no conflicts of interest.

Supplementary material

For supplementary material referred to in this article, please visit <https://doi.org/10.1017/S0007114521004955>

References

1. Mikkila V, Rasanen L, Raitakari OT, *et al.* (2005) Consistent dietary patterns identified from childhood to adulthood: the cardiovascular risk in young Finns study. *Br J Nutr* **93**, 923–931.
2. Ward KA, Prentice A, Kuh DL, *et al.* (2016) Life course dietary patterns and bone health in later life in a British birth cohort study. *J Bone Miner Res* **31**, 1167–1176.
3. van den Hooven EH, Ambrosini GL, Huang RC, *et al.* (2015) Identification of a dietary pattern prospectively associated with bone mass in Australian young adults. *Am J Clin Nutr* **102**, 1035–1043.
4. Pastorino S, Richards M, Pierce M, *et al.* (2016) A high-fat, high-glycaemic index, low-fibre dietary pattern is prospectively associated with type 2 diabetes in a British birth cohort. *Br J Nutr* **115**, 1632–1642.
5. Johns DJ, Lindroos AK, Jebb SA, *et al.* (2014) Tracking of a dietary pattern and its components over 10-years in the severely obese. *PLoS One* **9**, e97457.
6. Maddock J, Ambrosini GL, Griffin JL, *et al.* (2019) A dietary pattern derived using B-vitamins and its relationship with vascular markers over the life course. *Clin Nutr* **38**, 1464–1473.
7. Appannah G, Pot GK, Oddy WH, *et al.* (2018) Determinants of a dietary pattern linked with greater metabolic risk and its tracking during adolescence. *J Hum Nutr Diet* **31**, 218–227.
8. Appannah G, Pot GK, Huang RC, *et al.* (2015) Identification of a dietary pattern associated with greater cardiometabolic risk in adolescence. *Nutr Metab Cardiovasc Dis* **25**, 643–650.
9. Ambrosini GL, Johns DJ, Northstone K, *et al.* (2015) Free sugars and total fat are important characteristics of a dietary pattern associated with adiposity across childhood and adolescence. *J Nutr* **146**, 778–784.
10. Ambrosini GL, Emmett PM, Northstone K, *et al.* (2014) Tracking a dietary pattern associated with increased adiposity in childhood and adolescence. *Obesity* **22**, 458–465.
11. Ambrosini GL, Emmett PM, Northstone K, *et al.* (2012) Identification of a dietary pattern prospectively associated with increased adiposity during childhood and adolescence. *Int J Obes* **36**, 1299–1305.
12. Ambrosini GL, Oddy WH, Huang RC, *et al.* (2013) Prospective associations between sugar-sweetened beverage intakes and cardiometabolic risk factors in adolescents. *Am J Clin Nutr* **98**, 327–334.
13. Batis C, Sotres-Alvarez D, Gordon-Larsen P, *et al.* (2014) Longitudinal analysis of dietary patterns in Chinese adults from 1991 to 2009. *Br J Nutr* **111**, 1441–1451.

14. Da Mota Santana J, Alves de Oliveira Queiroz V, Monteiro Brito S, *et al.* (2015) Food consumption patterns during pregnancy: a longitudinal study in a region of the north east of Brazil. *Nutr Hosp* **32**, 130–138.
15. Gasser CE, Kerr JA, Mensah FK, *et al.* (2017) Stability and change in dietary scores and patterns across six waves of the longitudinal study of Australian children. *Br J Nutr* **117**, 1137–1150.
16. Gasser CE, Mensah FK, Kerr JA, *et al.* (2017) Early life socioeconomic determinants of dietary score and pattern trajectories across six waves of the longitudinal study of Australian children. *J Epidemiol Community Health* **71**, 1152–1160.
17. Hassannejad R, Kazemi I, Sadeghi M, *et al.* (2018) Longitudinal association of metabolic syndrome and dietary patterns: a 13-year prospective population-based cohort study. *Nutr Metab Cardiovasc Dis* **28**, 352–360.
18. Mishra B, McNaughton S, Bramwell G, *et al.* (2006) Longitudinal changes in dietary patterns during adult life. *Br J Nutr* **96**, 735–744.
19. Luque V, Escibano J, Closa-Monasterolo R, *et al.* (2018) Unhealthy dietary patterns established in infancy track to mid-childhood: the EU childhood obesity project. *J Nutr* **148**, 752–759.
20. O'Sullivan TA, Bremner AP, Mori TA, *et al.* (2016) Regular fat and reduced fat dairy products show similar associations with markers of adolescent cardiometabolic health. *Nutrients* **8**, 22.
21. Oellingrath IM, Svendsen MV & Brantsaeter AL (2011) Tracking of eating patterns and overweight – a follow-up study of Norwegian schoolchildren from middle childhood to early adolescence. *Nutr J* **10**, 106.
22. Wall CR, Thompson JM, Robinson E, *et al.* (2013) Dietary patterns of children at 3.5 and 7 years of age: a New Zealand birth cohort study. *Acta Paediatr* **102**, 137–142.
23. Northstone K & Emmett PM (2008) Are dietary patterns stable throughout early and mid-childhood? A birth cohort study. *Br J Nutr* **100**, 1069–1076.
24. Walthouwer MJ, Oenema A, Soetens K, *et al.* (2014) Are clusters of dietary patterns and cluster membership stable over time? Results of a longitudinal cluster analysis study. *Appetite* **82**, 154–159.
25. Northstone K, Smith AD, Newby PK, *et al.* (2013) Longitudinal comparisons of dietary patterns derived by cluster analysis in 7- to 13-year-old children. *Br J Nutr* **109**, 2050–2058.
26. Harrington JM, Dahly DL, Fitzgerald AP, *et al.* (2014) Capturing changes in dietary patterns among older adults: a latent class analysis of an ageing Irish cohort. *Public Health Nutr* **17**, 2674–2686.
27. Maddock J, Ziauddeen N, Ambrosini GL, *et al.* (2018) Adherence to a dietary approaches to stop hypertension (DASH)-type diet over the life course and associated vascular function: a study based on the MRC 1946 British birth cohort. *Br J Nutr* **119**, 581–589.
28. Martins MCT, Jaceldo-Siegl K, Orlich M, *et al.* (2017) A new approach to assess lifetime dietary patterns finds lower consumption of animal foods with aging in a longitudinal analysis of a health-oriented Adventist population. *Nutrients* **9**, 118.
29. Wang Y, Bentley ME, Zhai F, *et al.* (2002) Tracking of dietary intake patterns of Chinese from childhood to adolescence over a 6-year follow-up period. *J Nutr* **132**, 430–438.
30. Mikkila V, Rasanen L, Raitakari OT, *et al.* (2007) Major dietary patterns and cardiovascular risk factors from childhood to adulthood: the cardiovascular risk in young Finns study. *Br J Nutr* **98**, 218–225.
31. Lock EF, Hoadley KA, Marron JS, *et al.* (2013) Joint and individual variation explained (jive) for integrated analysis of multiple data types. *Ann Appl Stat* **7**, 523–542.
32. Chua SWY (1999) Iron and vitamin A nutrition to young Auckland children: an investigation into the methods to assess the nutritional status of micronutrients in 6–24 month olds. MSc Thesis, Massey University.
33. Ministry of Health (2003) *New Zealand Food: New Zealand Children. Key Results of the 2002 National Children's Nutrition Survey*. Wellington: Ministry of Health.
34. Freeman JV, Cole TJ, Chinn S, *et al.* (1995) Cross sectional stature and weight reference curves for the UK, 1990. *Arch Dis Child* **73**, 17–24.
35. Lurbe E, Cifkova R, Cruickshank JK, *et al.* (2009) Management of high blood pressure in children and adolescents: recommendations of the European society of hypertension. *J Hypertens* **27**, 1719–1742.
36. Team RC (2020) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
37. O'Connell MJ & Lock EF (2017) r.jive: Perform JIVE Decomposition for Multi-Source Data. <https://CRAN.R-project.org/package=r.jive> (accessed December 2020).
38. Wei T & Simko V (2017) R Package “Corrplot”: Visualization of a Correlation Matrix. <https://CRAN.R-project.org/package=corrplot> (accessed December 2020).
39. Xie Y (2020) knitr: A General-Purpose Package for Dynamic Report Generation in R. <https://CRAN.R-project.org/package=knitr> (accessed December 2020).
40. Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
41. Zhu H, Li G & Lock EF (2020) Generalized integrative principal component analysis for multi-type data with block-wise missing structure. *Biostatistics* **21**, 302–318.