Pimpin, L., Jebb, S. A., Johnson, L., Llewellyn, C., & Ambrosini, G. L. (2018). Sources and pattern of protein intake and risk of overweight or obesity in young UK twins. *British Journal of Nutrition*, *120*(7), 820-829. https://doi.org/10.1017/S0007114518002052

Sources and pattern of protein intake and risk of overweight or obesity in young UK twins

Authors: Laura Pimpin[1], Susan A Jebb[1, 2], Laura Johnson[3], Clare Llewellyn[4], Gina L Ambrosini[1,5*].

[1] Medical Research Council Human Nutrition Research, Cambridge, United Kingdom; [2] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom; [3] Centre for Exercise, Nutrition and Health Sciences, School for Policy Studies, University of Bristol, Bristol, United Kingdom; [4] Cancer Research UK Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, London, United Kingdom; [5]School of Population and Global Health, The University of Western Australia, Perth, Western Australia

Pimpin
Jebb
Johnson
Llewellyn
Wardle
Ambrosini

[*]**To whom correspondence should be addressed**: Gina Ambrosini, School of Population and Global Health, The University of Western Australia, Perth, Western Australia, email: Gina.Ambrosini@uwa.edu.au, phone: +61 8 6488 7375

Running head: Protein type & BMI, weight to 5 years

Keywords: dairy protein, BMI, weight, child

Abstract

High protein intake in young children is associated with excess gains in weight and body fat, but the specific role of different protein sources has yet to be described. The study objective was to investigate the role of different types of protein in the post weaning stage on weight, BMI, and overweight/obesity at 60 months. Intakes of animal, dairy and plant protein and a dietary pattern characterising the variation in protein types at 21 months of age were estimated using a 3-day diet diary in a cohort of 2154 twins; weight and height were recorded every three months from birth to 60 months. Longitudinal mixed effect models investigated associations between sources of protein intake or dietary pattern scores and BMI, weight, and overweight/obesity from 21 months up to 60 months. Adjusting for confounders, dairy protein intake at 21 months was positively associated with greater weight (46 g (95% CI 21;71) and BMI up to 60 months (0.04 kg/m^2 (95%CI 0.004;0.070)) and the odds of overweight/obesity at 3 years (OR 1.12(95% CI 1.00;1.24)). Milk showed associations of similar magnitude. A dietary pattern low in dairy and high in plant protein was associated with lower weight gain up to age 60 months, but not overweight/obesity. Intake of dairy products in early childhood is most strongly associated with weight gain, compared to other protein sources. A dietary pattern characterised by lower protein intake and greater diversity in protein sources at 2 years of age may confer a lower risk of excess weight gain.

Introduction

Recent investigations have identified that high protein consumption in young children may be associated with excess weight gain (1; 2; 3). In the UK, the National Diet and Nutrition Survey (NDNS) and the 2011 Diet and Nutrition Survey in Infants and Young Children show that children in the complementary feeding phase were consuming protein in excess of their requirements (4). In the Gemini cohort, the largest contemporary dietary dataset in the UK for children of approximately 2 years of age, children aged 21 months consumed on average 40g of protein per day, compared to the 15g/d recommended by the Department of Health for children aged 1 to 3 years (5).

Protein intake is a known dietary determinant of circulating Insulin-like Growth Factor-1 (IGF-1) levels in humans, (6) an important factor in linear growth in childhood (7), but also in adipocyte differentiation and maturation through direct induction of cellular multiplication (8), and through feedback inhibition of the lipolytic effect of growth hormone. Intakes of total, animal and dairy protein (from dairy and meat sources), but not vegetable protein (from cereal, fruit and vegetable sources), have been independently correlated with serum IGF-1 concentration, in observational studies of young children (2; 9). Trials of milk supplementation in both infants (9 to 12 m) and young children (6 to 8 y) have resulted in higher IGF-1 levels (10; 11). These results suggest that different sources of protein may have varying metabolic effects with consequences for growth in the early years.

Studies that have sought to identify associations between different protein types and excess adiposity in children, have led to mixed findings: In the Second Icelandic Infant study total animal protein, but not vegetable protein at 12 months, was associated with a higher BMI at six years. However, dairy protein and meat protein were not individual predictors of BMI at six years(2). A small Dutch cohort of 120 children also found significant associations between higher animal protein intake and increased BMI at eight years and greater odds of overweight and obesity at nine years (12). But it is unclear whether specific sources of animal protein, can explain the association between animal protein and BMI in this study. Total protein, total animal protein (including dairy and meat protein) at 12 months, but not at 18-24 months, showed the same positive association with BMI and adiposity at seven years, in a study of 203 German children (13). Analysing the sources of protein at 12 months separately, found only dairy protein, but not protein from meat or cereals, was associated with BMI at seven years. Interpretations of these studies are limited by their small sample sizes, variation in ages at exposure and outcome measurement, duration of follow-up and single point outcome measurement. In addition, these studies were often not comparable in their definition of protein sources; for example, some combine dairy with other animal protein into one single exposure variable.

Establishing which types of proteins are associated with weight gain and the dietary patterns by which they are consumed will help to both better understand their relationships and to translate this information into public health advice. Defining diets as a pattern of foods through analytic techniques such as reduced rank regression (RRR) helps explain the variation in specific nutrients of interest (e.g. protein) and makes it easier to interpret the dietary exposures at a food level. This approach takes into account hypothesised mechanisms such as the potential role of the protein-IGF-1-growth hormone axis (6; 7), or alterations of methyl donor metabolism and epigenetic processes (14) to allow clearer detection of the association with weight increases over the early childhood period.

The aims of this study were to: i) examine the prospective associations between specific protein sources, weight and obesity risk; ii) derive a prevalent dietary pattern that explains intakes of specific protein sources; and iii) examine the prospective associations between this dietary pattern, weight gain and obesity, in a cohort of young UK children.

Methods

Study population and design

The population consisted of 4804 twins enrolled in the Gemini study, a large population-based birth cohort of British families with twins born in England and Wales in 2007. Details of this cohort have been described previously (15). In brief, all families with live twin births during March to December 2007 (n=6754) were invited to participate by the Office for National Statistics. A total of 2402 families (35.6% response rate) were recruited and completed a baseline questionnaire when the children were approximately 8.2 (SD=2.2) months old, collecting information on: child anthropometrics; activity behaviours; birth complications; medical conditions; appetite traits and the introduction of solid foods; as well as parental feeding style, demographics, anthropometrics, health behaviours, parent illnesses and medical conditions. Data collection included dietary data collected at an average of 21 months of age (SD = 1.2) and repeated measures of anthropometrics from birth to 60 months of age (60 months), see Figure 1. The University College London Committee for the Ethics of Non-National Health Service Human Research granted ethical approval for the Gemini study.

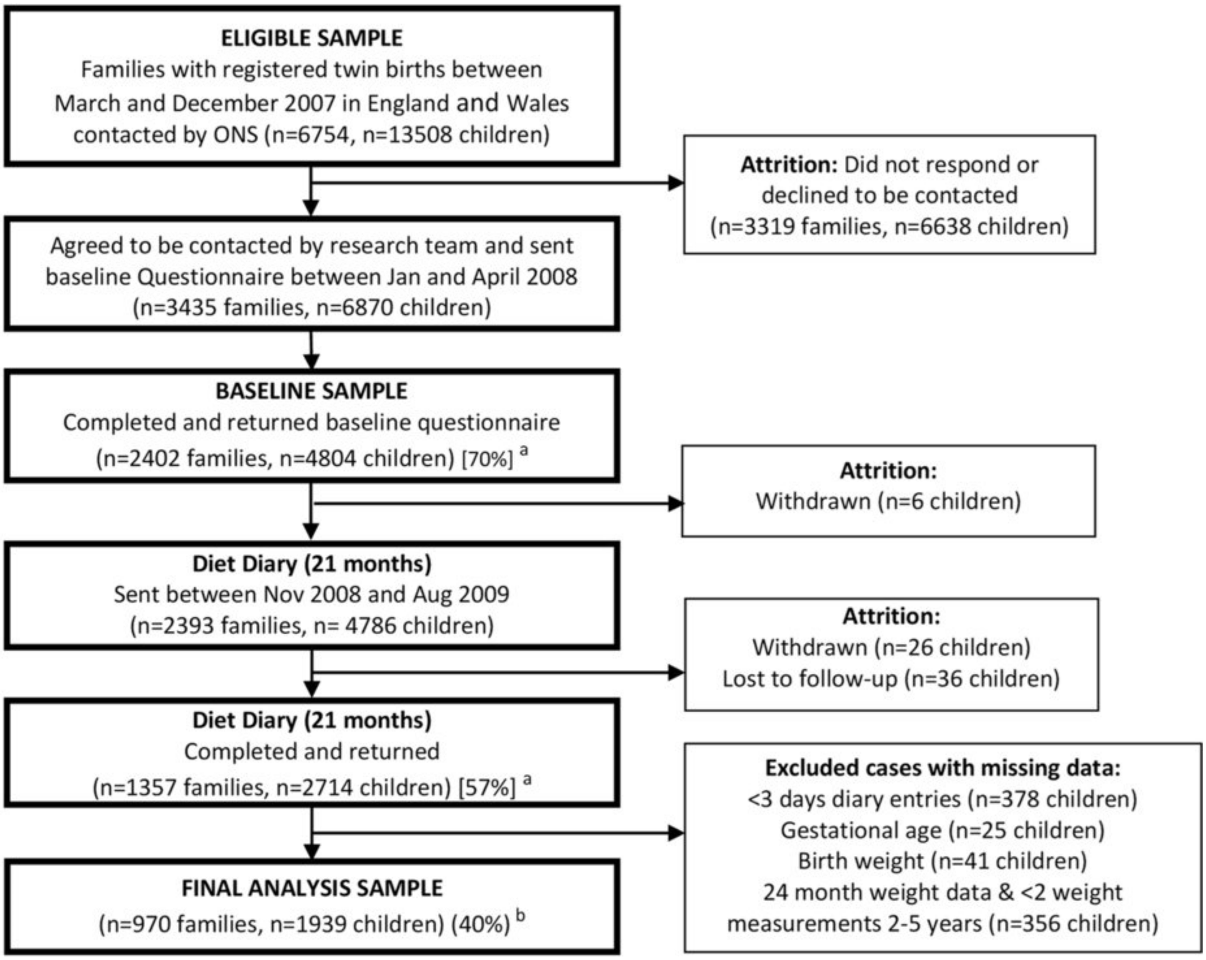

Figure 1. **Flowchart of Gemini study participants**

Dietary variables

Three-day diet diaries were sent to all Gemini families for completion between November 2008 and August 2009. Parents were provided with detailed instructions and portion guides adapted from the pre-school food atlas, to facilitate accurate estimation of intakes of different types of foods. Parents were asked to record all food and drinks consumed by each twin over three days (including any two weekdays and one weekend day). Completed diet diaries for 2714 children (56% of twins originally recruited) were received. The average age when the 3-day diet diary was completed was 21 months (SD=1.2). Diaries including only one recorded day, completed for children outside of the 18 to 27 months age range, or completed with more than 28 days between the first and last day of diary entry were excluded to ensure the data represented dietary intake for a given month of age. Diaries from twins of unknown zygosity were excluded. These exclusion criteria combined resulted in including 1216 diaries for a total of 2432 twins for analysis (50.6% of the original cohort). Despite instructions, some diaries were not completed using a combination of weekdays and weekend day (n=684), but these were included as diet for this age group is unlikely to be greatly biased by the day of the week reported, as it is thought to be for adults (16).

The estimated three-day diet diary method is similar to that used in the UK NDNS (17), coded using a continually updated food composition database (18) based on McCance and Widdowson's Composition of Foods 6th edition (19), related supplemental material and manufacturer information.

To quantify the consumption of protein more accurately, composite foods and meals were disaggregated (20). Portion sizes of discrete protein based-foods were based on reference portion sizes (21).

Individual foods were classified into 45 food groups reflecting broad food groups used in NDNS, with further disaggregation of groups (e.g. milks) to allow animal, plant, dairy protein sources to be separated (see supplementary information). Animal protein was defined as protein in g/d and as a proportion of total energy (%E) sourced from red and white meat, processed meats, fish, and eggs; but excluded dairy protein, as this was a specific protein source of interest. Using %E as a measure of protein intake reduces measurement error, as a consequence of the correlation between errors in macronutrients and energy as the total energy intake is the sum of the energy from the individual macronutrients (22). Use of self-reported energy intake energy adjustment of other self-reported dietary constituents has been proposed to also improve risk estimation in studies of diet-health associations. Subar et al, suggest that while self- reported energy intake as a measure of true energy intake may not be valid, it is valid to use self-reported energy intake for energy adjustment of other self-reported dietary constituents to improve risk estimation in studies of diet-health association (23).

No distinction was made between processed and unprocessed meat. Milk protein constituted a significant part of total dairy protein, and was separated from other dairy products such as cheese, yoghurts, butter and cream, to characterise intake of milk protein only. Formula milk and breast milk protein were classified under dairy protein. Although their composition differs from cow or goat's milk, the source of protein is closest to dairy and few consumers with low consumption levels in this cohort mean that this difference was unlikely to affect results strongly. Protein intake from cereals, fruit, vegetable, nuts, pulses and potatoes and plant-based milks were combined as plant protein.

Dietary pattern analysis

RRR was used to identify a combination of food intakes constituting a dietary pattern that could explain the maximum variation in intake of animal, dairy and plant protein (24). The RRR model included all 45 food groups (g/d) as predictors and the three intermediate variables of interest: animal protein, dairy protein and plant protein (all %E). As three intermediate variables were included, the RRR model identified three dietary patterns. The dietary pattern of most relevance was identified according to how much of the total variation in intermediate variables it explained. The RRR model scored each study participant for each dietary pattern with a z-score, indicating the degree to which their reported dietary intake reflected the observed dietary pattern, relative to others in the cohort.

Anthropometric variables

Methods for data collection of anthropometric and dietary variables have been described elsewhere (5; 25). Briefly, parental reports of child weight and height were collected every 3 months by electronic and postal questionnaires, from birth to 60 months of age, along with the date of each measurement. This was collected up to 24 months of age from healthcare professionals' records, after which parents were sent electronic weighing scales and height measurement wall charts along with detailed instructions on how to measure, record, and report their twins' weights and height every 3 months. Measurements were recorded and reported to researchers up to a median age of 55.0 (interquartile range (IQR): 42.0–60.7) months for the last recorded measurement at the time of this analysis. A total of 2154 twins with included diet diaries provided at least two anthropometric measurements up to 60 months. BMI was calculated as kg/m^2 and age specific International Obesity Task Force (IOTF) cut offs used to provide data on subjects' overweight/obesity status, and to allow results to be comparable with other studies conducted in other countries (26). Data on overweight/obesity status between 33 and 39 months and between 57 and 63 months were combined to provide data on overweight/obesity at 36±3 months and 60±3 months, respectively, to maximise data at these two specific time points. Children were also coded as ever or never being overweight or obese from diet diary to end of follow-up (at 60 months/5 years of age).

Statistical analyses

Protein intakes were analysed as absolute amounts (g/d) and as a proportion of total energy intake (% E), to investigate the role of total protein intake as well as the contribution of protein to the overall macronutrient composition of the diet. Associations between animal protein, plant protein and dairy protein intake at 21 months and repeated measures of weight and BMI were analysed using linear mixed-effect models. These used repeated measurements of BMI and weight from the first measurement available after the diet diary (median age 24.1 months (IQR: 22.1, 24.8)) up to 60 months of age as the outcome variable, time at measurement (months) as the level-1 predictor and measures of protein intake as the level 2 predictor variable. The best fitting models (according to likelihood ratio testing) included a random intercept and slope and unstructured covariance between the random effects at both the twin pair and the individual level. To adjust for the within-pair clustering of twins, all regression models included a cluster term for family.

Logistic mixed effect models, with the same random effects structure as described above, evaluated the prospective association between specific sources of protein intake at 21 months and the odds of being overweight or obese at 36 ± 3 months, 60± 3 months, and at any point up to 60± 3 months of age.

Dietary pattern z-scores at 21 months of age were investigated in relation to repeated measures of weight and BMI up to 5 years of age using the linear regression mixed effects models, and odds of overweight and obesity using logistic mixed effects models, as above.

Covariates were tested for inclusion in both the linear and logistic regression models using forward stepwise selection. These included factors identified in the literature as having a relationship with dietary intake and adiposity: demographic variables for gender, age at diet diary entry (months), zygosity, ethnicity (white vs. other) and family socio-economic status (SES, three categories). We also tested variables indicating prior nutrition and growth: feeding method in the first 3 months of life (seven categories ranging from exclusively breast-fed to exclusively bottle-fed), mean daily energy intake (kJ), maternal BMI at baseline (four categories for underweight, normal weight, overweight and obese), rate of prior weight gain between birth and time of dietary intake measurement (derived using mixed effect regression (27)) and birth weight (kg). Height (cm) was included in weight models only. Total intake of fat was additionally included in the model to allow interpretations of macronutrient substitution. In models where protein intake was analysed as %E, mean daily fat as a proportion of energy intake (%Efat) was included, while models of protein intake in g/d included total g/d fat intake.

After covariate testing, the adjusted models included exact age at dietary assessment, gender, birth weight, rate of previous growth (in weight gain), fat intake (%Efat or g/d), total energy intake and height (where weight was outcome). As no interaction between %E from total protein and age was detected in our previous analyses (27), no interaction was included these models.

Results

Dietary information was available for 2,432 children. The number of participants who provided weight and height up to 60 months varied according to the outcome measure. The prevalence of overweight and obesity was 12% at 36 months, 6 % at 60 months and 16% at any time point between 21 and 60 months (data not shown).

Intake of types of protein

Dairy sources represented almost half of all protein intake (Table 1 and Supplementary Table 2) of which three-quarters was from milk. Animal protein (defined as protein derived from meat, chicken, fish and eggs) and plant protein combined accounted for most of the remaining protein intake. Animal and dairy protein intake were modestly negatively inversely correlated in this population (r -0.24), and there was a strong inverse correlation between intakes of dairy and plant protein (r -0.54). There was no correlation between animal and plant protein (r -0.04) (data not shown).

Association between intake of types of protein and weight, BMI or odds of overweight/obesity.

Table 2 shows that intake of dairy protein and milk protein in g/d and as a proportion of total energy intake were both associated with greater weight and higher BMI up to 60 months. A one percent higher %E from dairy protein at 21 months was associated with 46 g (95% CI 21; 71) greater weight and with 0.04 (95% CI 0.004; 0.070) greater BMI on average, at any age between 21 and 60 months, after adjusting for confounders. Similarly, a one gram/d higher dairy protein intake was positively associated with greater weight and BMI on average up to 60 months (Table 2). Intake of milk protein showed slightly stronger associations with weight and BMI. The difference in milk protein intake between children in the 10th and 90th centile in the Gemini study was 15.1g/d (equivalent to approximately 415 ml of semi-skimmed milk). Applying the coefficient to this difference equates to a 300 g difference in weight and 0.35 BMI units, between children in the 10th and 90th percentile of milk protein intake.

These results were similar when examining the odds of overweight or obesity. In adjusted models, intake of dairy protein (in both %E and g/d) at the expense of energy from other protein or carbohydrates at 21 months, was weakly associated with increased odds of overweight or obesity at 3 years, but not up to or at 60 months of age (Table 3). The association for milk protein was of a similar magnitude but more consistent. A higher milk protein (g/d or %E) at the expense of other protein and carbohydrates was associated with increased odds of overweight or obesity at 36 and 60 months (Table 3).

No association was observed between animal protein or plant protein and weight, BMI or the odds of overweight or obesity at any age (Tables 2 & 3).

Dietary patterns of protein type

The three dietary patterns identified through the RRR analysis describe combinations of foods consumed that explain protein intakes: Dietary pattern 1 was modestly correlated with %E from animal protein (r = 0.23), had a strong negative correlation with %E from dairy protein (r = -0.75) and a strong positive correlation with % E from plant protein (r = 0.62). Factor loadings (Figure 2) show that this “low-dairy, high-plant protein” pattern was characterised by low intakes (negative factor loadings) of animal milk (largely cow’s milk), cream and cheese, chocolate, yoghurts and dairy desserts, and positively associated with intakes of plant-based milks (e.g. soy milk), pulses, fruit, bread, vegetables, potatoes, pasta, rice, white meat (positive factor loadings). The other two patterns explained very small amounts of total variation in protein sources (17% and 7%) and were therefore less meaningful. Table 4 shows the proportion of total variation in protein sources explained by each of the three dietary patterns identified. Pattern 1 explained the most variation in all three protein source variables (41%). As the first pattern explained the most variation in all protein sources, we took this pattern forward for analysis in relation to anthropometric measurements. The z-scores for this dietary pattern had a mean of 0 (SD 0.96) and ranged from -3.6 to 5.6.

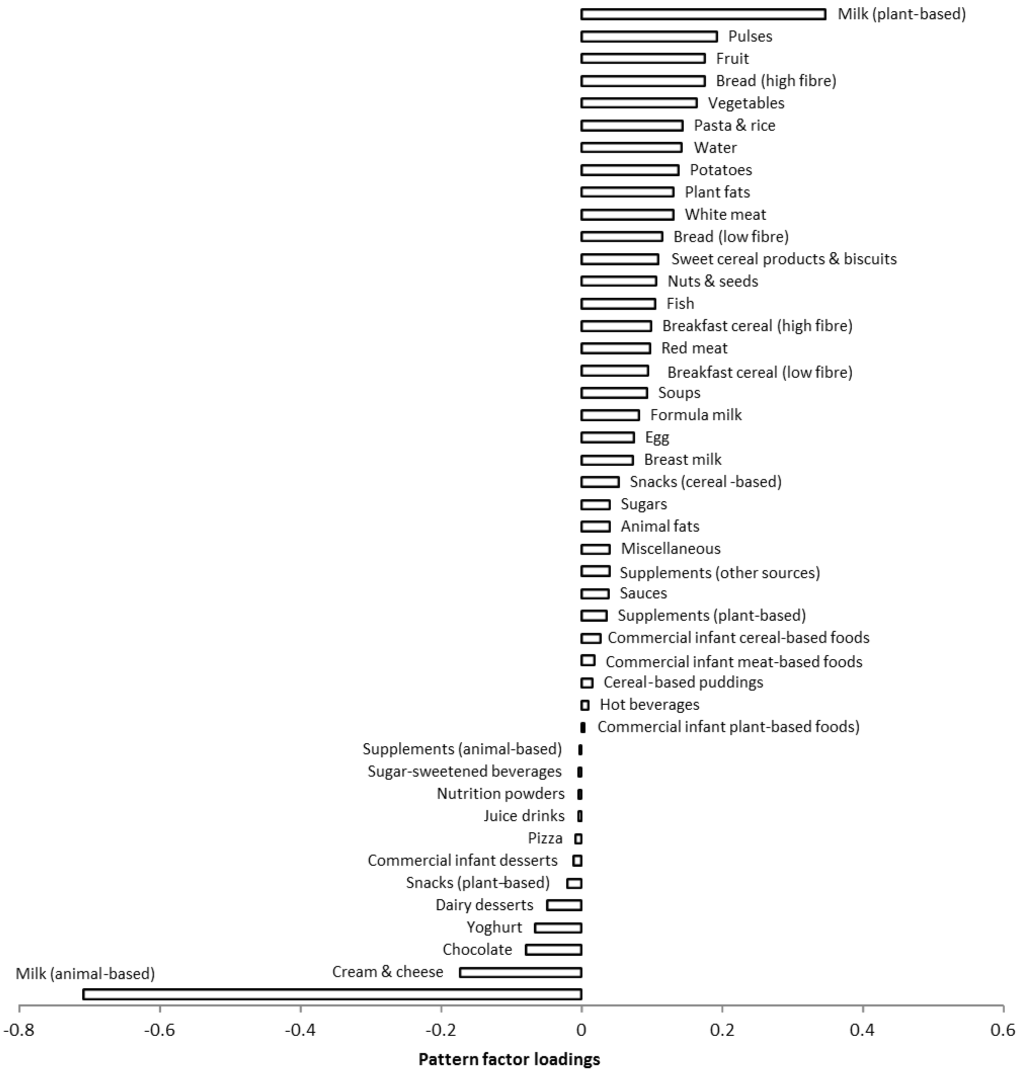

Figure 2. Food group factor loadings for the 'low dairy' dietary pattern (Characterisations of the food groups in supplementary data.)

Association between 'low -dairy', high-plant protein" dietary pattern z-score and weight, BMI or odds of overweight/obesity.

The 'low-dairy, high-plant protein' dietary pattern was negatively associated with weight up to 60 months (Table 2).: A 1 SD higher z-score for this dietary pattern at 21 months of age was associated with 50 g lower weight (95% CI 6; 107g) at any age between 21 and 60 months. There was a significant association between z-score for this dietary pattern and lower odds of being overweight or obese at 36 months (OR 0.78 (CI 0.63; 0.98) but not at 60 months, and some evidence of an association at any time point between 21 and 60 months

(OR 0.84 (95% 0.70; 1.00). Alternatively, a 1 SD unit lower z-score for this dietary pattern at 21 months, indicating a high dairy protein and low plant protein intake, would be associated with 1.22 (95% 1.02; 1.37) greater odds of being overweight or obese at 36 months of age.

Discussion

This analysis shows that the association between total protein intake at 21 months of age and later weight gain is largely explained by dairy protein (especially milk), which showed the strongest and most consistent associations with weight (adjusted for height) and BMI, compared to other protein types.

We identified a low dairy protein, high plant protein dietary pattern prospectively associated with lower weight gain, BMI and lower risk of overweight and obesity. This pattern was strongly characterised by low intakes of animal milk (and to a lesser extent other dairy products) and high intakes of plant-based milks and a range of food groups representing plant and animal protein sources. The factor loading for animal-based milk was three times that of other food groups, which corroborates our finding that milk intake at this age explains a protein-based dietary pattern predictive of greater weight gain and increased obesity risk.

A low score for this low -dairy, high-plant protein dietary pattern (and subsequently increased risk of excess weight gain), could be achieved by a high consumption of animal-based milk (as indicated by the strongest factor loading) as well as high intakes of other dairy products, and a low consumption of fruits, vegetables, pulses, bread, potatoes, pasta, rice, breakfast cereals, white meat, fish and red meat. Such a dietary pattern at 21 months of age is indicative of a diet reflecting incomplete weaning, whereby milk remains an integral part of dietary intake, potentially substituting food-based meals or supplementing the energy intake in children who are already consuming other meals. Collectively, these results suggest that children who are consuming cow’s milk as their predominant source of protein at the expense of other sources of protein and food-based meals by 18 to 27 months of age, are at a greater risk of excess weight gain, while those with a diet characterised by lower intakes of milk are at lower risk of excess weight gain.

About 300ml of milk (just over half a pint or 310 grams) will provide the 350mg of calcium a day required by children 1 to 3 years of age for British children (28) and is below the threshold associated with a risk of iron deficiency, as has been showed in children who consume a large amount of milk (>500ml per day) (29). The median intake for animal milk for children of this age in the Gemini sample was 375 g (IQR: 246; 490), but children above the 75th centile for milk intake had intakes averaging 610 g/d (SD 113), up to a maximum intake of 1091 g/d. Milk intakes between 490 g/d and 1091 g/d would provide between 322 kcal/d and 720 kcal/d, or 30-70% of the average energy requirements in this population. Our findings are supported by other analyses of the Gemini dataset, which found that the cohort greatly exceeded the recommended levels of calcium, another marker of milk intake, with a mean intake of 842 mg (SD 4.8) representing 240% of the recommended nutrient intake for children aged 1-3 years (5). Furthermore, an analysis by MacDonald et al of sleep and night-time energy intake in the Gemini Study highlighted that excess energy intake at night-time was primarily milk (30). Milk was hypothesised to be provided by parents as an aid to encourage children to sleep through the night rather than for nutritive purposes. This offers one insight into how milk may be supplementing other food consumption. However, the factor loadings for meal-based foods in our dietary pattern also suggest that milk may be displacing energy from foods, as well as supplementing it.

Comparison with other studies

The Second Icelandic Infant Nutrition Study reported a significant relationship between animal protein intake, but not plant protein intake, and BMI at six years of age (2). There was no specific association for meat or dairy protein alone, although the sample size of 199 children in this study potentially limited the detection of effects from smaller subgroups of intake. Quintile analyses also revealed a difference between consumers in the top and bottom quartiles of animal protein in the Icelandic cohort. Similarly, the Generation R study of 2911 children found an association between dietary protein intake measured at 1 year (using a food frequency questionnaire and BMI SDS and body fatness SDS measured at 6 years. However, this association was only significant for animal protein, and not vegetable protein (RR 0.06 BMI SDS (CI 0.01; 0.1) per 10g animal protein intake per day. This study did not investigate the specific role of dairy however, as animal protein included both dairy and non-dairy sources.(31)

To our knowledge, this is the first study to extract a dietary pattern based on different protein sources using RRR in a population of this age. The dietary pattern observed in this study is different to that of other studies of obesity in older children, which focus more strongly on fibre, energy density and fat intakes (32). It does not preclude the possibility that dietary patterns evolve as children age, perhaps shifting from a milk-protein dominated pattern to an energy-dense, low fibre pattern through childhood, but further research is required to study changing dietary patterns throughout childhood and adolescence.

Strengths and limitations

The largest contemporary dietary dataset for toddlers in the UK combined with repeated measures of weight, make this study well powered to detect small associations. The dietary pattern approach in this study corroborates the previous analysis of individual protein sources. Moreover, it provides important information about food sources of protein that are linked to greater or lower adiposity risk in young children, and offers insights into which foods may be displacing others. We used RRR to derive dietary patterns that can be used to test specific hypothesis on pathways from diet to development of a disease. The effects of the whole diet in the form of dietary patterns can help improve the interpretation of the protein source analysis into evidence that can be more easily translated into food based advice, for instance milk intake, which is the single most important food group contributing to protein intake in children of this age. However, the pattern observed in this study explained variations in protein type only, not other nutrients. Accordingly, it does not necessarily represent the most common dietary pattern in UK children of this age, but rather the pattern of foods that best explain animal, plant and dairy protein intakes in this cohort of young children.

Using dietary data collected at one time point is a limitation of this study, as it may not capture dietary variation. However, there is evidence that dietary patterns are determined early in life, and track in both infancy (33) and older childhood (34). The average age at dietary intake 21 months and the use of repeated measurement for anthropometrics, means that the window between exposure and outcome will vary across individuals in the cohort.

Misreporting of food consumption is a potential limitation to all self-reported dietary assessment methods, although younger ages are associated with lower rates of dietary under-reporting (35). However, we have used dietary records, the gold standard in self-reported dietary assessment, as shown by comparisons with doubly-labelled water to assess energy intakes (36). Furthermore, in a recent biomarker study in 450 postmenopausal women found that a food frequency questionnaire, a 4-day diet record, and three 24-hour recalls each explained respectively: 3.8%, 7.8%, and 2.8% of biomarker variation for energy; 8.4%, 22.6%, and 16.2% of biomarker variation for protein; and 6.5%, 11.0%, and 7.0% of biomarker variation for protein density(37). There is potential for under-estimating food wastage, resulting in over-reporting of actual consumption in this age group (38; 39). While the gold standard procedure for diet diaries involves weighing intakes, a validation study by Lanigan et al. using 72 children aged six to 24 months found no evidence that estimated food records are less accurate than weighed records for assessing energy and nutrient sub-class intakes in this age group (40). The nutrient disaggregation method may also be source of possible errors due to data entry and misclassification, although automation of the process avoided this to some extent. Interestingly, this study's intake estimates for animal, dairy and plant protein were similar to that described by Thorisdottir et al. (2).The analysis may be confounded by interactions with other nutrients which were not included. The calcium content of dairy products has been hypothesised to have a negative impact on body composition and growth, while fibre, could play a role in the observed differences in effects seen between different types of protein. However, analyses were adjusted for total energy intake and total fat intake to account for different composition of foods providing protein from different sources, as well as measurement error (41). Obesity rates at 60 months of age in this study were lower than the national average of 22% in children aged four to five in the 2012-2013 National Child Measurement Programme (42). Results from the smaller Health Survey England 2012 found a 23-28% prevalence of overweight and obesity in children aged two to 10 between 2010 and 2012 (43). This may be due to the younger age of the Gemini children, compared to these national estimates, but also the smaller birth weight of twins. The over-representation of children from higher SES and white ethnicity families may also contribute to the slightly lower rates of overweight and obesity.(27) The 16% of children in the Gemini sample who were classified as overweight or obese at any time point from time of diet diary to follow-up may be a more accurate representation of overweight and obesity status during the follow-up period, , which greatly reduces sample size, and therefore may lead to further limitation. Indeed, the analysis of dietary intake on odds of obesity at 3 and 5 years is based on a much smaller sample size, than longitudinal analyses which can use repeated data from individuals measured several times, but at any time point in follow-up. With a decrease in sample size, particular in analysis individuals with BMI data at 60 months, comes loss of power, which may go some way to explaining the stronger association with dairy and milk protein in younger ages (36 months).

Conclusion

Total protein intakes in this cohort of young UK children are high and are largely sourced from milk. A higher dairy and milk protein intake is prospectively associated with greater weight gain and higher risk of overweight or obesity in this cohort. A more diverse dietary pattern with less emphasis on dairy protein, particularly animal-based milk, and greater intakes of plant and animal foods appears to be associated with more appropriate weight gain. This analysis provides important information on food groups to target in intervention, education and recommendations to prevent excessive weight gain in early life. Future research must consider the quality and source of protein in addition to the absolute amount. Parents and carers may need more detailed advice about appropriate weaning practices to avoid over reliance on milk and reduce the risk of excessive weight gain.

Abbreviations: BMI: body mass index; IQR: Interquartile range; MRC HNR: Medical Research Council Human Nutrition Unit; NDNS: National diet and nutrition survey; NHS: National Health Service; PCA; Principal component analysis RRR: Reduced rank regression; SES: Socio-economic status; OR: odds ratio; SD: standard deviation;

Acknowledgements

The authors would like to acknowledge the substantial intellectual contribution by Professor Jane Wardle who sadly passed away prior to publication.

Financial support

Supported by a program grant from the UK Medical Research Council (U105960389 to SAJ) and MRC PhD studentship (LP). The Gemini study was supported by a programme grant from Cancer Research UK (C1418/ A7974 to JW). Diet diary coding was funded by consultancy fees from Danone Baby Nutrition (to JW and LJ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. SAJ is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust and the Oxford NIHR Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflict of Interest

JW received grants from Cancer Research UK and from Danone Baby Nutrition during the conduct of the study. JW & LJ received institutional consultancy fees from Danone Baby Nutrition from 2009-2012All other authors declared no conflicts of interest.

Authorship

The authors’ responsibilities were as follows—LP, SJ, and GLA: designed the research (developed and conceived of the study); JW was principal investigator of the Gemini study with responsibility for data collection; CL coordinated the collection of demographic and anthropometric data; LJ designed and co-ordinated the collection and coding of dietary data; LP undertook the food disaggregation, performed statistical analyses and drafted the manuscript; all authors advised on the analyses or interpretation of data and contributed to manuscript preparation.

References

1. Gruszfeld D, Weber M, Gradowska K *et al.* (2016) Association of early protein intake and pre-peritoneal fat at five years of age: Follow-up of a randomized clinical trial. *Nutrition, metabolism, and cardiovascular diseases : NMCD* **26**, 824-832.
2. Thorisdottir B, Gunnarsdottir I, Palsson GI *et al.* (2014) Animal protein intake at 12 months is associated with growth factors at the age of six. *Acta Paediatrica* **103**, 512517.
3. Voortman T, van den Hooven EH, Tielemans MJ *et al.* (2016) Protein intake in early childhood and cardiometabolic health at school age: the Generation R Study. *European journal of nutrition* **55**, 2117-2127.
4. Public Health England (2014) *National Diet and Nutrition Survey: Results from Years 1-4 (combined) of the Rolling Programme (2008/2009 – 2011/12)*. London.
5. Syrad H, Llewellyn CH, van Jaarsveld CH *et al.* (2016) Energy and nutrient intakes of young children in the UK: findings from the Gemini twin cohort. *The British journal of nutrition* **115**, 1843-1850.
6. Martin RM, Holly JM, Gunnell D (2011) Milk and linear growth: programming of the IGF-I axis and implication for health in adulthood. *Nestle Nutrition Workshop Series Paediatric Programme* **67**, 79-97.
7. Rogers I, Emmett P, Gunnell D *et al.* (2006) Milk as a food for growth? The insulin-like growth factors link. *Public Health Nutr* **9**, 359-368.
8. Wabitsch M, Hauner H, Heinze E *et al.* (1995) The role of growth hormone/insulin-like growth factors in adipocyte differentiation. *Metab* **44**, 45-49.
9. Hoppe C, Udam TR, Lauritzen L *et al.* (2004) Animal protein intake, serum insulin-like growth factor I, and growth in healthy 2.5-y-old Danish children. *Am J Clin Nutr* **80**, 447-452.
10. Larnkjaer A, Hoppe C, Mølgaard C *et al.* (2009) The effects of whole milk and infant formula on growth and IGF-I in late infancy. *Eur J Clin Nutr* **63**, 956-963.
11. Hoppe C, Molgaard C, Juul A *et al.* (2004) High intakes of skimmed milk, but not meat, increase serum IGF-I and IGFBP-3 in eight-year-old boys. *Eur J Clin Nutr* **58**, 1211-1216.
12. Weijs PJ, Kool LM, van Baar NM *et al.* (2011) High beverage sugar as well as high animal protein intake at infancy may increase overweight risk at 8 years: a prospective longitudinal pilot study. *Nutr J* **10**, 95.
13. Günther A, Remer T, Kroke A *et al.* (2007) Early protein intake and later obesity risk: which protein sources at which time points throughout infancy and childhood are important for body mass index and body fat percentage at 7 y of age? *Am J Clin Nutr* **86**, 1765-1772.
14. Lind MV, Larnkjær A, Mølgaard C *et al.* (2017) Dietary protein intake and quality in early life: impact on growth and obesity. *Curr Opin Clin Nutr Metab Care* **20**, 71-76.
15. van Jaarsveld CH, Johnson L, Llewellyn C *et al.* (2010) Gemini: a UK twin birth cohort with a focus on early childhood weight trajectories, appetite and the family environment. *Twin research and human genetics : the official journal of the International Society for Twin Studies* **13**, 72-78.
16. Thane CW, Stephen AM (2006) Day-to-day variation in food and nutrient intakes of British adults. *Public Health Nutr* **9**, 102.
17. Bates B, Lennox A, Prentice A *et al.* (2012) *National Diet and Nutrition Survey: Headline results from Years 1, 2 and 3 (combined) of the Rolling Programme (2008/2009-2010/11). A survey carried out on behalf of the Department of Health and Food Standards Agency*. FSA, DH, NatCen, UCL, HNR.
18. Fitt E, Cole D, Ziauddeen N *et al.* (2014) DINO (Diet In Nutrients Out) - an integrated dietary assessment system. *Public Health Nutr*, 1-8.
19. Food Standards Agency (2002) *McCance and Widdowson's The Composition of Foods*. sixth summary ed. Cambridge: The Royal Society of Chemistry.
20. Fitt E, Mak TN, Stephen AM *et al.* (2010) Disaggregating composite food codes in the UK National Diet and Nutrition Survey food composition databank. *Eur J Clin Nutr* **64 Suppl 3**, S32-36.
21. Food Standards Agency (editor) (1994) *Food Portion Sizes (Maff Handbook).* London.
22. Kipnis V, Subar AF, Midthune D *et al.* (2003) Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14-21.

23. Subar AF, Freedman LS, Tooze JA *et al.* (2015) Addressing current criticism regarding the value of self-report dietary data. *J Nutr* **145**, 2639-2645.
24. Hoffmann K, Schulze MB, Schienkiewitz A *et al.* (2004) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* **159**, 935-944.
25. van Jaarsveld CH, Boniface D, Llewellyn CH *et al.* (2014) Appetite and growth: a longitudinal sibling analysis. *JAMA pediatrics* **168**, 345-350.
26. Cole TJ, Lobstein T (2012) Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric obesity* **7**, 284-294.
27. Pimpin L, Jebb S, Johnson L *et al.* (2016) Dietary protein intake is associated with body mass index and weight up to 5 y of age in a prospective cohort of twins. *Am J Clin Nutr* **103**, 389-397.
28. Department of Health NHS choices - Milk and dairy foods. http://www.nhs.uk/livewell/goodfood/pages/milk-dairy-foods.aspx
29. Maguire JL, Lebovic G, Kandasamy S *et al.* (2013) The relationship between cow's milk and stores of vitamin D and iron in early childhood. *Pediatrics* **131**, e144-e151.
30. McDonald L, Wardle J, Llewellyn CH *et al.* (2015) Sleep and nighttime energy consumption in early childhood: a population-based cohort study. *Pediatric obesity* **10**, 454-460.
31. Voortman T, Braun K, Kiefte-de Jong J *et al.* (2016) Protein intake in early childhood and body composition at the age of 6 years: the Generation R Study. *Int J Obes* **40**.
32. Ambrosini GL (2014) Childhood dietary patterns and later obesity: a review of the evidence. *Proc Nutr Soc* **73**, 137-146.
33. Lioret S, Betoko A, Forhan A *et al.* (2015) Dietary patterns track from infancy to preschool age: cross-sectional and longitudinal perspectives. *J Nutr* **145**, 775-782.
34. Ambrosini GL, Emmett PM, Northstone K *et al.* (2014) Tracking a dietary pattern associated with increased adiposity in childhood and adolescence. *Obes* **22**, 458-465.
35. Lioret S, Touvier M, Balin M *et al.* (2011) Characteristics of energy under-reporting in children and adolescents. *The British journal of nutrition* **105**, 1671-1680.
36. Freedman LS, Commins JM, Moler JE *et al.* (2014) Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* **180**, 172-188.
37. Prentice RL, Mossavar-Rahmani Y, Huang Y *et al.* (2011) Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Am J Epidemiol* **174**, 591-603.
38. Lennox A, Sommerville, J, Ong, K, Henderson, H, Allen, R (2013) *Diet and Nutrition Survey of Infants and Young Children, 2011.* . Department of Health.
39. Forrestal S (2011) Energy intake misreporting among children and adolescents: a literature review. *Matern Child Nutr* **7**, 112-127.
40. Lanigan J, Wells J, Lawson M *et al.* (2001) Validation of food diary method for assessment of dietary energy and macronutrient intake in infants and children aged 6-24 months. *Eur J Clin Nutr* **55**, 124-129.
41. Willett WC, Howe GR, Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65**, 1220S-1228S; discussion 1229S-1231S.
42. Public Health England (2013) *National Child Measurement Programme - England, 2012-13 school year*.
43. The NHS Information Centre (2013) *Health Survey for England 2012.* London.

Tables

Table 1 - Intake of protein by source (g/d, % total protein and % of total energy)

N=2432	Grams per day Mean (SD) range	Proportion of total protein intake (%) Mean (SD)	Proportion of total energy intake (%) Mean (SD)
Total Protein	39.9 (8.6) 12;76	100	15.7
Total animal protein	9.2 (5.0) 0; 41.5	22 (10.0)	3.6
Total dairy protein	18.6 (6.6) 0; 43.9	46 (12.0)	7.3
Total milk protein	13.2 (6.0) 0; 37.7	33 (12.6)	5.2
Total plant protein	12.1 (3.5) 0; 41.5	32 (10.0)	4.8

Animal protein defined as protein derived from meat, chicken, fish and eggs, dairy protein defined as protein derived from milk, cheese and yoghurt products.

Table 2 - Association between intake of animal, dairy and plant protein (%E and g/d) and repeated measures of weight and BMI between 21 and 60 months.

	Weight (kg) up to 60 months; Model 1 N= 2154 /Model 2 N = 2050			**BMI (kg/m^2) up to 60 months; Model 1 N= 2154 /Model 2 N = 1769**		
	β	**95% CI**	**p-value**	**β**	**95% CI**	**p-value**
Animal protein (%total energy)						
Model 1	0.016	(-0.017; 0.049)	0.344	0.004	(-0.029; 0.037)	0.814
Model 2 (adjusted)	0.020	(-0.004; 0.044)	0.123	0.004	(-0.031; 0.039)	0.818
Animal protein (g/d)						
Model 1	0.009	(-0.003; 0.021)	0.144	0.004	(-0.010; 0.018)	0.528
Model 2 (adjusted)	0.009	(-0.001; 0.019)	0.073	0.003	(-0.011; 0.017)	0.652
Dairy protein (%total energy)						
Model	0.036	(0.009; 0.063)	**0.009**	0.031	(0.002; 0.060)	**0.037**
Model 2 (adjusted)	0.046	(0.021; 0.071)	**<0.001**	0.037	(0.004; 0.070)	**0.034**
Dairy protein (g/d)						
Model 1	0.018	(0.008; 0.028)	**0.001**	0.014	(0.002; 0.026)	**0.017**
Model 2 (adjusted)	0.017	(0.007; 0.027)	**<0.001**	0.013	(-0.001; 0.027)	**0.050**
Milk protein (%total energy)						
Model 1	0.046	(0.018; 0.073)	**<0.001**	0.054	(0.025; 0.083)	**<0.001**
Model 2 (adjusted)	0.046	(0.022; 0.069)	**<0.001**	0.065	(0.030; 0.100)	**<0.001**
Milk protein (g/d)						
Model 1	0.021	(0.009; 0.032)	**<0.001**	0.023	(0.011; 0.034)	**<0.001**
Model 2 (adjusted)	0.020	(0.010; 0.029)	**<0.001**	0.023	(0.009; 0.037)	**0.001**

Plant protein (%total energy)						
Model 1	-0.015	(-0.064; 0.034)	0.547	0.007	(-0.044; 0.058)	0.779
Model 2 (adjusted)	0.005	(-0.034; 0.044)	0.825	0.024	(-0.041; 0.089)	0.470
Plant protein (g/d)						
Model 1	0.006	(-0.014; 0.026)	0.590	0.009	(-0.013; 0.031)	0.389
Model 2 (adjusted)	0.000	(-0.018; 0.018)	0.984	0.008	(-0.019; 0.035)	0.556
Low dairy-high plant protein						
Dietary pattern z-score*						
Model 1*	-0.038	(-0.101; 0.025)	0.238	-0.054	-0.121; 0.013	0.115
Model 2 (adjusted) ***	-0.050	(-0.107; -0.006)	**0.008**	-0.065	-0.139; 0.008	0.081

Animal protein defined as protein derived from meat, chicken, fish and eggs, dairy protein defined as protein derived from milk, cheese and yoghurt products. β per 1-unit increase in dietary variable, linear mixed effects model (p value is for test of null hypothesis: β=0); Model 1 includes protein exposure variable of interest and repeated measures of weight (kg) or BMI between the first measurement available after diet diary and the last measurement up to 60 months as the outcome, adjusted for total energy intake (kJ) and height (cm, weight models only); *Model 2 includes all variables from Model 1 and adjusts for gender, age at diet diary (months), birth weight (kg), modelled rate of prior weight gain, or BMI, depending on the response variable, from birth to time of diet diary, and total fat intake (%Efat or g/d), depending on measure of protein modelled; Estimates in bold are p-values <0.05 (Bonferroni correction for multiple testing indicates a p-value of <0.0001 to reject the null hypothesis)

Table 3 - Association between intake of animal, dairy and plant protein (%E and g/d) and odds of overweight and obesity during follow-up to 60 ± 3 months of age

	Overweight or obese at 36 ± 3 months Model 1 N= 1385 Models 2 = 1159		**Overweight or obese at 60 ± 3 months Model 1 N= 1058 Models 2 = 855**		**Ever overweight or obese up to 60±3 months Model 1 =1854 Model 2 =1534**	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Animal protein (%total energy)						
Model 1	1.03 (0.93; 1.13)	0.624	0.94 (0.80; 1.09)	0.396	0.97 (0.90; 1.04)	0.394
Model 2 (adjusted)	1.04 (0.94; 1.15)	0.448	0.94 (0.80; 1.10)	0.447	0.99 (0.91; 1.09)	0.877
Animal protein (g/d)						
Model 1	1.06 (0.97; 1.05)	0.421	0.97 (0.91; 1.02)	0.318	1.00 (0.97; 1.03)	0.886
Model 2 (adjusted)	1.02 (0.98; 1.07)	0.240	0.97 (0.92; 1.03)	0.412	1.00 (0.96; 1.03)	0.902
Dairy protein (%total energy)						
Model 1	1.13 (1.03; 1.23)	**0.010**	0.98 (0.85; 1.13)	0.807	1.07 (0.99; 1.15)	0.085
Model 2 (adjusted)	1.12 (1.00; 1.24)	**0.048**	0.95 (0.83; 1.09)	0.445	1.07 (0.98; 1.16)	0.094
Dairy protein (g/d)						
Model 1	1.05 (1.01; 1.09)	**0.007**	0.99 (0.94; 1.10)	0.790	1.03 (1.00; 1.05)	**0.027**
Model 2 (adjusted)	1.05 (1.00; 1.09)	**0.033**	1.01 (0.93; 1.11)	0.772	1.02 (0.99; 1.05)	0.145
Milk protein (%total energy)						
Model 1	1.14 (1.05; 1.24)	**0.002**	0.99 (0.88; 1.12)	0.917	1.08 (1.03; 1.18)	**0.003**

Model 2 (adjusted)	1.14 (1.04; 1.25)	**0.005**	0.98 (0.87; 1.12)	0.842	1.12 (1.04; 1.22)	**0.004**
Milk protein (g/d)						
Model 1	1.05 (1.02; 1.08)	**0.002**	0.99 (0.95; 1.04)	0.868	1.04 (1.01; 1.07)	**0.007**
Model 2 (adjusted)	1.05 (1.01; 1.09)	**0.007**	0.99 (0.95; 1.05)	0.836	1.04 (1.01; 1.07)	**0.014**
Plant protein (%total energy)						
Model 1	0.88 (0.70; 1.04)	0.112	1.03 (0.78; 1.34)	0.857	0.95 (0.94; 1.08)	0.426
Model 2 (adjusted)	0.84 (0.67; 1.04)	0.115	1.01 (0.74; 1.38)	0.964	0.93 (0.80; 1.09)	0.393
Plant protein (g/d)						
Model 1	0.94 (0.88; 1.02)	0.123	1.01 (0.90; 1.13)	0.889	1.00 (0.96; 1.05)	0.838
Model 2 (adjusted)	0.97 (0.90; 1.04)	0.390	1.02 (0.92; 1.14)	0.684	0.97 (0.91; 1.03)	0.366
Low dairy- high plant protein						
Dietary pattern z-score						
Model 1	0.79 (0.65; 0.97)	**0.025**	0.94 (0.70; 1.24)	0.640	0.86 (0.73; 1.01)	0.059
Model 2 (adjusted)	0.78 (0.63; 0.98)	**0.034**	0.92 (0.67; 1.26)	0.585	0.84 (0.70; 1.00)	0.054

Animal protein defined as protein derived from meat, chicken, fish and eggs, dairy protein defined as protein derived from milk, cheese and yoghurt products. OR per 1-unit increase in dietary variable, logistic mixed effects model (p value is for test of null hypothesis: OR=1); Model 1 adjusts for total energy intake (kJ); Model 2 additionally adjusted for age at diet diary (months), gender, birth weight (kg), modelled rate of prior weight gain and total fat intake (%Efat or g/d, depending on measure of protein modelled); Estimates in bold are p-values <0.05; (Bonferroni correction for multiple testing indicates a p-value of <0.0001 to reject the null hypothesis)

Table 4 - Proportion of variation in intermediate variables explained by each dietary pattern extracted

	Intermediate variables included in RRR			
	%E Animal Protein	**%E Dairy Protein**	**%E Plant Protein**	**Total variation in intermediate variables combined***
Dietary pattern 1	7	70	47	41
Dietary pattern 2	41	0	9	17
Dietary pattern 3	10	9	9	7

* variation of all three intermediate variables is not an additive process so is not a sum of the variation for each variable separately.

Dietary pattern 1 represents high intake of plant-based milks (mostly soy milk), pulses, fruits and vegetables and high fibre bread and with a low reported intake of milk from dairy sources, dairy products and chocolate. As such it is can be referred to as a 'low dairy' dietary pattern.

Dietary pattern 2 represents: High intakes of animal protein, from meat and fish, but very little intake of dairy.

Dietary pattern 3 represents: A mixed intake of all protein types, with no clearly discernible combination of foods consumed (only 7% of total variable in source of %E protein explained by this pattern).

Animal protein defined as protein derived from meat, chicken, fish and eggs, dairy protein defined as protein derived from milk, cheese and yoghurt products.

Supplementary information

Supplementary Table 1: Foods included in the food groups from the Gemini study three-day diet diaries

45 Food groups	Included foods
Milk (animal-based)	skimmed, semi-skimmed and whole cow's milk, other animal-based milk, milk-based flavoured drinks
Milk (plant-based)	plant-based milk
Water	water
Juice drinks	fruit-based drinks, baby/infant processed juice drinks
Formula milk	all formula milks
Hot beverages	powdered beverages, coffee, tea
Sugar-sweetened beverages	carbonated soft drinks
Breast milk	breast milk
Pizza	pizza
Pasta & rice	pasta and pasta dishes, rice and rice dishes
Cereal (low fibre)	other cereals and dishes
Cereal (high fibre)	oat-based cereals
Fruit	fresh, dried, canned and cooked fruit
Yoghurt	yoghurt, drinking yoghurts
Cream & cheese	cream, fromage frais, cheese
Dairy desserts (& milk-based puddings)	ice cream, dairy desserts & milk-based puddings*
Vegetables	tomatoes, brassicacea, yellow, red and dark green leafy vegetables, other vegetables,
Pulses	pulses, lentils, baked beans
Red meat	beef and veal and dishes, lamb and dishes, red meat, venison, bacon and ham, processed pies, other processed meats, sausages and burgers and kebab, liver and dishes, other offal and dishes
White meat	pork and dishes, chicken and turkey and dishes, other game birds
Fish	white fish, oily fish, shellfish
Bread (low fibre)	white, other breads, crisp breads
Bread (high fibre)	wholemeal, brown, granary, wheat germ
Animal fats	butter, animal based fats
Plant fats	oils, plant based fats
Potatoes	potatoes, potato products
Sweet cereal products & biscuits	biscuits, pastries, buns, pies, cereal bars
Cereal-based puddings (not milk)	cereal-based puddings (not milk)
Chocolate	chocolate-based products
Sugars	jam, marmalade, chutney, pickles, pure sugars, other sugars including syrups, honey, sugar-based products, sorbets, lollies
Snacks (cereal-based)	savoury biscuits, crackers, cereal-based snacks
Snacks (vegetable-based)	potato-based snacks, vegetable-based snacks
Nuts & seeds	nuts & seeds

Commercial infant vegetable-based foods	vegetable-based manufactured ready meals
Commercial infant meat-based foods	meat-based manufactured ready meals
Commercial infant cereal-based foods	dried cereals
Commercial infant desserts	fruit only purees, biscuits
Eggs	egg and egg dishes.
Supplements (vegetal-based)	supplements (vegetal)
Supplements (animal-based)	supplements (animal)
Supplements (other sources)	supplements (other)
Nutrition powders	nutrition powders
Soups	soups
Sauces	dressings and mayonnaise, cooking sauces (gravies, pesto, brown sauce, soy sauce ketchup)
Miscellaneous	dried herbs & spices, salt and artificial sweeteners

Supplementary Table 2 Intake (g/d), (kJ/d) and proportion of total daily energy intake from 45 food groups in the whole diet diary sample and in consumers only

				Consumers only			Whole sample (including non-consumers)		
Total food consumed (Median (IQR))		**Consumers in sample**		**Total intake (g/d)**	**Total energy intake (kJ/d)**	**Proportion of total daily energy intake (%)**	**Total food intake (g/d)**	**Total energy intake (kJ/d)**	**Proportion of total daily energy intake (%)**
Liquids		**N**	**(%)**						
Milk	Milk (animal-based)	2340	96	375 (246;490)	998 (651;1320)	23 (16;30)	367 (224;485)	978 (592;1306)	23 (14;30)
	Milk (plant-based)	72	3	125 (51;213)	228 (108;428)	6 (3;10)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Formula milk		335	14	275 (196;392)	804 (567;1140)	19 (13;27)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Breast milk		41	2	200 (100;300)	578 (289;867)	11 (7;18)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Water		2382	98	231 (148; 359)	0 (0; 0)	0 (0; 0)	228 (142; 353)	0 (0; 0)	0 (0; 0)
Juice drinks		1037	43	70 (28;125)	115 (41;204)	3 (1;5)	0 (0;53)	0 (0;86)	0 (0;2)
Other beverages	Hot beverages	236	10	30 (7; 110)	9 (2; 94)	0 (0; 2)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Sugar-sweetened beverages	1071	44	32 (16; 83)	9 (4; 30)	0 (0; 1)	0 (0; 26)	0 (0; 6)	0 (0; 0)
Solids									
Cereal products	Pizza	221	9	65 (40;95)	693 (420;1043)	15 (10;23)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Pasta & rice	1870	77	56 (38;91)	266 (165;407)	6 (4;9)	44 (10;78)	200 (49;348)	5 (1;8)
	Cereal (low fibre)	1840	76	15 (9;26)	222 (134;339)	5 (3;8)	11 (1;21)	164 (15;293)	4 (0;7)
	Cereal (high fibre)	1873	77	20 (18;30)	297 (260;446)	8 (6;10)	20 (8;27)	297 (114;386)	7 (3;9)
Fruit		2378	98	119 (80;163)	381 (253;514)	9 (6;12)	117 (77;162)	374 (240;510)	9 (6;12)
Dairy	Yoghurt	1282	53	83 (60;100)	284 (247;427)	7 (5;10)	33 (0;85)	127 (0;284)	3 (0;7)
	Cream & cheese	2187	90	39 (19;66)	323 (213;457)	8 (5;11)	34 (14;63)	297 (173;433)	7 (4;10)
	Dairy desserts (& milk-based puddings)	1094	45	60 (35;90)	297 (218;452)	7 (5;10)	0 (0;57)	0 (0;278)	0 (0;7)
Vegetables	Vegetables	2339	96	55 (33;80)	77 (46;125)	2 (1;3)	52 (31;78)	75 (42;121)	2 (1;3)

Supplementary Table 2 Intake (g/d), (kJ/d) and proportion of total daily energy intake from 45 food groups in the whole diet diary sample and in consumers only

				Consumers only			Whole sample (including non-consumers)		
Total food consumed (Median (IQR))		**Consumers in sample**		**Total intake (g/d)**	**Total energy intake (kJ/d)**	**Proportion of total daily energy intake (%)**	**Total food intake (g/d)**	**Total energy intake (kJ/d)**	**Proportion of total daily energy intake (%)**
	Pulses	1230	51	43 (26;73)	168 (110;294)	4 (2;7)	4 (0;43)	17 (0;170)	0 (0;4)
Meat & fish	Red meat	2030	83	32 (21;52)	285 (179;460)	7 (4;11)	27 (14;47)	242 (104;417)	6 (3;9)
	White meat	1410	56	28 (15;45)	188 (101;334)	4 (3;8)	11 (0;30)	74 (0;223)	2 (0;5)
	Fish	1446	60	38 (24;56)	265 (148;469)	6 (4;10)	18 (0;42)	104 (0;327)	3 (0;7)
Bread	Bread (low fibre)	1815	75	30 (18;44)	333 (208;485)	8 (5;11)	21 (0;38)	242 (0;417)	6 (0;10)
	Bread (high fibre)	1569	65	31 (20;42)	322 (213;446)	7 (5;10)	19 (0;36)	197 (0;361)	5 (0;8)
Fats & oils	Animal fats	1136	47	7 (5;11)	214 (132;316)	5 (3;7)	0 (0;7)	0 (0;206)	0 (0;5)
	Plant fats	1963	81	5 (2; 9)	128 (67; 210)	3 (2; 5)	4 (1; 8)	98 (30; 187)	2 (1; 4)
Potato		2152	88	45 (31;63)	205 (132;299)	5 (3;7)	42 (23;61)	187 (98;283)	4 (2;7)
Sweet cereal-based products	Sweet cereal products & biscuits	2192	90	22 (14;36)	371 (242;574)	9 (6;13)	20 (11;34)	346 (197;546)	8 (5;12)
	Cereal-based puddings (not milk)	73	3	64 (43;85)	589 (371;785)	13 (8;17)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Added sugars and confectionery	Chocolate	849	35	13 (7; 18)	272 (144; 367)	6 (3; 8)	0 (0; 7)	0 (0; 153)	0 (0; 4)
	Sugars	1402	58	9 (5; 15)	104 (56; 170)	2 (1; 4)	3 (0; 10)	35 (0; 123)	1 (0; 3)
Savoury snacks	Snacks (cereal-based)	979	40	10 (6;17)	193 (121;346)	5 (3;8)	0 (0;8)	0 (0;153)	0 (0;4)
	Snacks (vegetable-based)	664	27	12 (8;17)	247 (181;353)	6 (4;8)	0 (0;5)	0 (0;108)	0 (0;2)
	Nuts & seeds	392	16	6 (2;13)	129 (23;281)	3 (1;6)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Commercial infant foods	Commercial infant vegetable-based foods	147	6	130 (95; 190)	356 (269; 531)	9 (7; 13)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Commercial infant meat-based foods	215	9	182 (125; 230)	514 (354; 612)	11 (8; 14)	0 (0; 0)	0 (0; 0)	0 (0; 0)

Supplementary Table 2 Intake (g/d), (kJ/d) and proportion of total daily energy intake from 45 food groups in the whole diet diary sample and in consumers only

				Consumers only			**Whole sample (including non-consumers)**		
Total food consumed (Median (IQR))		**Consumers in sample**		**Total intake (g/d)**	**Total energy intake (kJ/d)**	**Proportion of total daily energy intake (%)**	**Total food intake (g/d)**	**Total energy intake (kJ/d)**	**Proportion of total daily energy intake (%)**
	Commercial infant cereal-based foods	43	2	100 (38; 125)	254 (152; 449)	7 (4; 11)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Commercial infant desserts	362	15	100 (75; 115)	298 (170; 357)	7 (4; 9)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Egg	Egg	1026	42	42 (16;60)	299 (109;410)	6 (3;10)	0 (0;30)	0 (0;205)	0 (0;5)
Supplements	Supplements (animal-based)	88	4	1 (1; 1)	18 (18; 35)	0 (0; 1)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Supplements (vegetal-based)	2	0.1	5 (4; 5)	154 (140; 168)	3 (3; 3)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Supplements (other sources)	189	8	1 (1; 1)	0 (0; 0)	0 (0; 0)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Nutrition powders	6	0.2	76 (2; 196)	470 (67; 833)	12 (2; 22)	0 (0; 0)	0 (0; 0)	0 (0; 0)
Sauces, condiments and soups	Soups	319	13	66 (22; 125)	88 (14; 232)	2 (0; 5)	0 (0; 0)	0 (0; 0)	0 (0; 0)
	Sauces	1633	67	21 (8; 36)	59 (32; 124)	1 (1; 3)	9 (0; 28)	32 (0; 84)	1 (0; 2)
	Miscellaneous	1444	59	1 (0; 1)	8 (4; 14)	0 (0; 0)	0 (0; 1)	2 (0; 9)	0 (0; 0)