



## Major and trace mineral composition of milk from lactating women following vegan, vegetarian and omnivore diets

Maryanne T. Perrin<sup>1\*</sup>, Roman Pawlak<sup>2</sup>, Nicholas Judd<sup>3</sup>, Jessica Cooper<sup>1</sup> and George L. Donati<sup>3</sup>

<sup>1</sup>Department of Nutrition, University of North Carolina Greensboro, 319 College Ave, 318 Stone Building, Greensboro, NC 27412, USA

<sup>2</sup>Department of Nutrition Science, East Carolina University, Health Science Building, Greenville, NC 27858, USA

<sup>3</sup>Department of Chemistry, Wake Forest University, Salem Hall, Winston-Salem, North Carolina 27109, USA

(Submitted 12 July 2022 – Final revision received 3 December 2022 – Accepted 19 December 2022 – First published online 23 December 2022)

### Abstract

Approximately one-in-ten reproductive age adults in the USA follow a plant-based diet, yet there is limited information on the influence of vegan and vegetarian diets on the mineral composition of breast milk. This study explored the major and trace mineral composition in breast milk and associations with maternal diet patterns. We used a cross-sectional design to collect a single sample of breast milk from individuals following vegan ( $n$  23), vegetarian ( $n$  19) and omnivore ( $n$  21) diet patterns. Plant-based diet ( $n$  42) was defined as following either vegan or vegetarian diets. Sixteen minerals were assessed using inductively coupled plasma mass spectrometry and inductively coupled plasma optical emission spectrometry. Data were evaluated using traditional statistical techniques and five different machine learning approaches. The distribution of Se (median; quartile 1 and 3) was significantly different between groups (vegetarians 21, 18–26  $\mu\text{g/l}$ ; vegans 19, 18–25  $\mu\text{g/l}$  and omnivores 17, 14–20  $\mu\text{g/l}$ ;  $P = 0.007$ ) using a Kruskal–Wallis test. Machine learning techniques also identified Se as a potential biomarker for differentiating breast milk by maternal diet pattern. Individuals following a plant-based diet generally had a lower BMI, higher breast milk Se and lower breast milk I and Fe concentrations compared with those following omnivore diets. This suggests that maternal dietary pattern (plant-based *v.* omnivore) may be helpful clinical information to consider when caring for the breast-feeding dyad, with the strongest evidence related to differences in Se concentration.

**Key words:** Human milk: Breast milk: Plant-based diet: Selenium

A recent Gallup poll reported that 8% of Americans followed a vegetarian or vegan diet, with higher rates in reproductive age groups (10% among ages 18–29 and 12% in ages 30–49)<sup>(1)</sup>. Prior research suggests that adults following vegan or vegetarian diets may consume low amounts of some minerals including Ca, I, Se and Zn<sup>(2–6)</sup>. Among pregnant women, individuals following a vegetarian diet have significantly lower Zn intake and status compared with those following omnivore diets<sup>(7)</sup>. In an acknowledgment of these trends, the National Institute of Health has highlighted the need for more research to understand the impact of vegan and vegetarian diet patterns during lactation on the composition of human milk<sup>(8)</sup>.

Research regarding minerals in breast milk from vegan and vegetarian mothers is scarce, with only two studies identified in a recent review by Karcz *et al.*<sup>(9)</sup> A semi-longitudinal study by Finley *et al.* looked at well-nourished omnivore and vegetarian mothers ( $n$  52) from 1 to 20 months postpartum and reported

that breast milk Fe, K, Mg, Ca, Zn and Na were not influenced by maternal intake<sup>(10)</sup>. The authors noted that supplements contributed to the mineral intake of vegetarian mothers which could have explained some of the null findings. Debski *et al.* conducted a study of lacto-ovo-vegetarians ( $n$  26) and omnivore ( $n$  12) lactating women in California. While no difference in Se intake between the vegetarian and omnivore mothers was observed, there were significant differences in breast milk Se (22.2 (SD 0.8) ng/ml and 16.8 (SD 1.3), respectively;  $P < 0.01$ )<sup>(11)</sup>. It is important to note that the authors did not report the lactation stage which may have contributed to the findings, as Se has been observed to decrease in human milk over the first months postpartum<sup>(12)</sup>. With evidence of differing intake of some minerals when following a vegetarian or vegan diet that may translate into differences in breast milk composition, the purpose of this study was to explore relationships between vegan, vegetarian and omnivore diets during lactation and the concentration

**Abbreviations** ICP-MS, inductively coupled plasma mass spectrometry; ICP-OES, inductively coupled plasma optical emission spectrometry; RFI, random forest feature importance; SVM-RFE, support vector machine recursive feature elimination.

\* **Corresponding author:** Maryanne T. Perrin, email [mperrin@uncg.edu](mailto:mperrin@uncg.edu)

of 16 minerals in human milk, controlling for other factors including lactation stage (infant age) and the use of multi- or prenatal vitamins.

## Methods

### Study overview

We conducted a cross-sectional study that enrolled lactating women following vegan, vegetarian or omnivore dietary patterns. Se was selected to perform a power calculation based on reports of low intake among adults following a vegan *v.* an omnivore diet<sup>(3)</sup>. Using descriptive statistics for Se in breast milk of vegetarians *v.* non-vegetarians reported by Debski *et al.*, unbalanced groups of nineteen vegetarian and twenty-one omnivore diet and an  $\alpha$  of 0.05, this study had a power of 0.97 for detecting differences in breast milk Se<sup>(11)</sup>. Detailed methods for study recruitment and sample collection have previously been reported<sup>(13,14)</sup>. Briefly, participants were lactating individuals who resided in the USA and consented to provide a single breast milk sample and complete a dietary screener to determine their diet pattern. The dietary screener asked about the consumption frequency of five different food groups: meat (including beef, lamb, pork, poultry); dairy products (milk, cheese and yogurts); eggs; fish and *n*-3 containing margarine, as well as the use of multi-micronutrient or prenatal supplements. Frequency responses were never; rarely (less than 1 time per month); sometimes (1–4 times/month) and often (more than 4 times/month). Participants were classified as vegan if they never consumed meat and fish and consumed non-meat animal products (e.g. dairy products, eggs) less than monthly (never or rarely). Vegetarians were characterised as not eating meat but regularly consuming other animal-based products. When vegan and vegetarian were combined into a single group for analysis, this was classified as a plant-based diet. Exclusion criteria included being < 2 weeks postpartum, being pregnant and known health conditions that could influence metabolism or B-12 status (e.g. gene mutation of methylene tetrahydrofolate reductase, hypothyroidism, hyperthyroidism, celiac disease, liver disease). Milk samples were collected in the morning by complete expression from a single breast, stored in a breast milk storage bag and frozen until transport to the research lab.

### Sample analysis

Mineral concentrations in milk were determined by inductively coupled plasma mass spectrometry (ICP-MS) and inductively coupled plasma optical emission spectrometry (ICP-OES), using an Agilent 8800 ICP-MS/MS and an Agilent 5110 ICP-OES (Agilent Technologies). Helium or hydrogen gas, flowing at 3.5 or 4.0 ml/min, respectively, was used in the ICP-MS collision–reaction cell to minimise spectral interferences. Additional details on the operating conditions used in ICP-MS and ICP-OES determinations are summarised in online Supplementary Table S1. Milk samples were prepared according to a procedure adapted from Dubascoux *et al.*<sup>(15)</sup>, which consisted of adding 0.30 ml (0 ml for ICP-OES) of distilled–deionised water (18 M $\Omega$ -cm, Purelab Option-Q, Elga) and 4.50 ml (4.98 ml

for ICP-OES) of an aqueous solution (AlkS) containing 1% v/v ammonia (NH<sub>4</sub>OH, Suprapur™, Millipore Sigma), 1% v/v 2-propanol (Semiconductor grade, Alfa Aesar), 0.1% m/v ethylene diamine tetra-acetic acid (EDTA, Electrophoresis grade, Alfa Aesar) and 5 × 10<sup>−4</sup>% v/v Triton X-100 (Electrophoresis grade, Sigma Chemical Co.) to 0.20 ml (20.0  $\mu$ l for ICP-OES) of milk. Blank solutions contained 0.50 ml (20.0  $\mu$ l for ICP-OES) of distilled–deionised water and 4.50 ml (4.98 ml for ICP-OES) of AlkS. Calibration solutions were prepared with 9.0 ml (8.0 ml for ICP-OES) of AlkS and adequate volumes of standard reference solutions of As, Ca, Cd, Cr, Cu, Fe, I, K, Mg, Mn, Mo, Na, P, Pb, Se and Zn (1000 or 10 mg/l, High Purity Standards) and distilled–deionised water for a final volume of 10.0 ml. Ca, K, Mg, Na and P were determined by ICP-OES, while the other elements were determined by ICP-MS. The analytical method was validated by addition and recovery experiments employing two different human milk samples. Spike concentrations were chosen based on a preliminary semi-quantitative analysis of the samples. For As, Cd, I, Mn, Mo, Pb and Se, a concentration spike of 20  $\mu$ g/l was adopted, while 40  $\mu$ g/l was used for Cr, Cu, Fe and Zn. For Ca, K, Mg, Na and P, a spike concentration of 4 mg/l was chosen.

### Statistical analysis

A Shapiro–Wilk test was used to evaluate data normalcy, and a Levene test was used to evaluate homogeneity of variance. Only maternal age, K and Mg had normal distributions; all variables had homogeneous variance ( $P > 0.05$ ) except lactation stage and diet duration. We used traditional statistical methods, as well as five different machine learning techniques to explore relationships between breast milk mineral composition and maternal diet pattern. The use of machine learning techniques overcomes some limitations associated with small sample sizes and facilitates the identification of hidden patterns in the data. Considering the relatively small data set, we have used multiple statistical methods to confirm variables associated with different maternal diet patterns<sup>(16)</sup>. Descriptive statistics were used to describe mineral composition in the full data set; correlation coefficients were computed to look for bivariate relationships between minerals and categorical variables were evaluated using a Fisher's exact test. Statistical and machine learning techniques used to probe for differences between diet groups included: (i) Kruskal–Wallis analysis with a Dunn's test for multiple comparisons using Holm's procedure to adjust  $P$  values for non-parametric data and ANOVA with Tukey's test for normally distributed data; (ii) random forest feature importance (RFI), which is a supervised, model-based embedded feature selection method based on random forests and the Gini index to identify the most important variables to split the data into different classification groups (in this case maternal diet)<sup>(17,18)</sup>; (iii) Boruta, which is a supervised feature selection wrapper based on random forest classification and backward feature elimination that ranks variables responsible for classifying the study samples into the different diet groups<sup>(19)</sup>; (iv) ReliefF, which is a supervised feature selection filter based on the  $k$  nearest neighbours technique to identify variables responsible for distinguishing maternal diet patterns<sup>(20,21)</sup>; (v) support vector machine recursive feature

**Table 1.** Characterisation of study participants (Median and quartiles; mean values and standard deviations)

	Overall ( <i>n</i> 63)		By maternal diet pattern						<i>P</i>
			Vegan ( <i>n</i> 23)		Vegetarian ( <i>n</i> 19)		Omnivore ( <i>n</i> 21)		
	Median	Quartile 1, quartile 3	Median	Quartile 1, quartile 3	Median	Quartile 1, quartile 3	Median	Quartile 1, quartile 3	
Age (years)*									
Mean	32		33		33		32		0.81
SD	5		5		5		4		
BMI	22.5	21.1, 26.3	22.0 <sup>a</sup>	20.6, 23.6	22.4 <sup>a</sup>	21.1, 25.0	26.3 <sup>b</sup>	22.2, 28.0	0.022
Infant age (weeks)	28	17, 48	25 <sup>a</sup>	14, 46	50 <sup>a</sup>	20, 93	25 <sup>a</sup>	18, 34	0.031
Gravida	2	1, 3	2	1, 3	2	1, 3	2	2, 3	0.50
Diet Duration (years)	7	3, 25	5 <sup>a</sup>	2, 10	6 <sup>a</sup>	3, 9	30 <sup>b</sup>	25, 33	<0.001
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Race/Ethnicity† (%)									0.65
Asian	1	2	1	4	0	0	0	0	
White	53	84	19	83	17	90	17	81	
Hispanic	2	3	1	4	1	5	0	0	
Mixed	7	11	2	9	1	5	4	19	
Multi-micronutrient supplement use† (%)	41		17		10		14		0.38
%	65 %		73.9 %		52.6 %		66.7 %		

Data represent median (quartile 1, quartile 3) analysed using a Kruskal–Wallis test, unless otherwise noted. Values with a different superscript were significantly different using Dunn's test for multiple comparisons.

\* Data represent mean and standard deviation analysed using ANOVA.

† Data represent number (prevalence) of participants reporting use of a multi- or prenatal vitamin; differences evaluated using Fisher's exact test.

elimination (SVM-RFE), which ranks the different variables according to their importance for maternal diet classification and assigns a *P*-value (at the 95 % confidence level) to each of them<sup>(22,23)</sup> and (vi) penalised logistic regression with Lasso penalty, which shrinks the less contributive variables to zero and identifies the most relevant features within each diet group<sup>(24,25)</sup>.

Elements with more than 50 % of the samples presenting a concentration lower than the analytical method's limit of quantification were removed from the data set (i.e. As, Cd, Cr and Mo). For the remaining elements, based on Succop *et al.*, entries with concentration values below the limit of detection were kept as they were<sup>(26)</sup>. For entries with a value of zero, a random number between zero and the limit of detection was employed. The R packages used for machine learning analyses were caret and randomForest for RFI; Boruta; caret and CORElearn for ReliefF; sigFeature for SVM-RFE and glmnet for penalised logistic regression with Lasso penalty.

Statistical analysis was conducted using SAS 9.4 (SAS Software) and the R programming language<sup>(27)</sup> (R Foundation for Statistical Computing). This study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects/patients were approved by the the Institutional Review Boards at the University of North Carolina Greensboro and East Carolina University (16-0310 and 16-001726, respectively). Consent forms were provided to all participants via email, and written consent was waived by the Institutional Review Boards due to minimal risk of participation.

## Results

Seventy-four subjects completed the original study, and milk samples from sixty-three participants were available for mineral

analysis based on the following maternal diet classification: vegan (*n* 23), vegetarian (*n* 19) and omnivore (*n* 21). Information about study participants is summarised in Table 1. There were no differences by diet group in maternal age, gravida or use of a multi-micronutrient or prenatal supplements. There were differences (median, quartile 1 to quartile 3) in BMI (vegans 22.0, 20.6–23.6; vegetarian 22.4, 21.1–25.0; omnivores 26.3, 22.2–28.0; *P* = 0.022), infant age at the time of sample collection (vegans 25, 14–46 weeks; vegetarians 50, 20–93 weeks; omnivores 25, 18–34 weeks; *P* = 0.031) and diet duration (vegans 5, 2–10 years; vegetarian 6, 3–9 years; omnivore 30, 25–33 years; *P* < 0.001). Multi-micronutrient or prenatal supplement use was reported by most study participants (41/63; 65 %) but did not differ statistically by maternal diet pattern (17/23, 74 % for vegans; 10/19, 52 % for vegetarians; 14/21, 67 % for omnivores; *P* = 0.38).

## Overview of milk mineral composition

Both ICP-MS and ICP-OES methods applied to determine mineral concentrations in breast milk presented adequate accuracy. Analyte percent recoveries for all elements evaluated were in the 93–108 % range, with relatively low standard deviations (online Supplementary Table S2). Adequate limit of detection and limit of quantification values were also obtained for determinations in the original breast milk sample (i.e. before any dilution) for both analytical methods, with values in the 0.1–80 µg/l and 0.07–30 mg/l for ICP-MS and ICP-OES, respectively (online Supplementary Table S3). More than 50 % of the samples had mineral concentrations below the limit of quantification for arsenic, cadmium, chromium and molybdenum. Concentrations of these minerals and the number of samples above the limit of quantification were as follows: arsenic (1–11 µg/l; *n* 9); cadmium (1–4 µg/l; *n* 3); chromium (3–23 µg/l;

**Table 2.** Mineral concentrations in breast milk by maternal diet pattern (Median and quartiles; mean values and standard deviations)

Mineral (units)	By maternal diet pattern								
	Overall ( <i>n</i> 63)		Vegan ( <i>n</i> 23)		Vegetarian ( <i>n</i> 19)		Omnivore ( <i>n</i> 21)		<i>P</i>
	Median	Quartile 1, quartile 3	Median	Quartile 1, quartile 3	Median	Quartile 1, quartile 3	Median	Quartile 1, quartile 3	
Ca (mg/l)	223	195, 255	229	192, 244	223	206, 271	222	188, 255	0.90
Cu (µg/l)	121	66, 184	107	63, 184	123	62, 166	146	79, 232	0.50
I (µg/l)	80	55, 147	62	40, 132	80	63, 147	93	71, 179	0.10
Fe (µg/l)	159	130, 216	150	131, 186	147	126, 227	184	146, 216	0.42
Pb (µg/l)	1.0	0.1, 2.0	1.0	0.1, 2.0	1.0	0.1, 2.0	1.0	1.0, 3.0	0.22
Mg (mg/l)*									
Mean	32		32		34		31		0.55
SD	9		9		12		7		
Mn (µg/l)	3	2, 4	3	2, 4	3	2, 3	3	2, 4	0.88
P (mg/l)	122	110, 133	125	119, 139	120	110, 138	115	108, 126	0.16
K (mg/l)*									
Mean	416		425		407		416		0.55
SD	53		54		45		60		
Se (µg/l)	19	15, 24	19 <sup>a</sup>	18, 25	21 <sup>a</sup>	18, 26	17 <sup>b</sup>	14, 20	0.007
Na (mg/l)	88	66, 139	91	63, 141	93	69, 174	88	65, 113	0.41
Zn (µg/l)	681	311, 1284	1012	80, 1393	480	80, 835	681	382, 988	0.38

Data represent median (quartile 1, quartile 3) evaluated with a Kruskal–Wallis test unless otherwise noted. Values with a different superscript were significantly different using Dunn's test for multiple comparisons.

\* Data represent means and standard deviations evaluated with ANOVA test.

*n* 12) and molybdenum (1–20 µg/l; *n* 19). The final statistical and machine learning analysis included twelve minerals: Ca, Cu, Fe, I, K, Mg, Mn, Na, P, Pb, Se and Zn.

Descriptive statistics for minerals in breast milk for the entire study population are presented in Table 2. Infant age was weakly associated with Se ( $r=0.472$ ;  $P<0.001$ ) and Zn ( $r=-0.486$ ;  $P<0.001$ ). The minerals that had the least variability (min, max) were Ca (151–375 mg/l), K (270–548 mg/l) and P (75–190 mg/l). There was a greater than 10-fold difference in five minerals: Cu (1–3132 µg/l), Pb (0–88 µg/l), I (18–1719 µg/l), Mn (1–13 µg/l) and Zn (80–2312 µg/l). Significant correlations between minerals were mostly moderate ( $r=0.4-0.69$ )<sup>(28)</sup> and included Pb and Mn ( $r=0.732$ ;  $P<0.001$ ); Se and Na ( $r=0.629$ ;  $P<0.001$ ); Fe and Mn ( $r=0.545$ ;  $P<0.001$ ); Fe and P ( $r=0.543$ ;  $P<0.001$ ); P and Se ( $r=0.502$ ;  $P<0.001$ ).

#### Relationship with maternal diet using multiple statistical techniques

Table 2 and Fig. 1 provide descriptive statistics and visual plots by maternal diet pattern. The only mineral with significant differences among the diet groups was Se (Kruskal–Wallis,  $P=0.007$ ). Using Dunn's method for pair-wise comparisons, median (quartile 1, quartile 3) Se composition was significantly higher ( $P=0.008$  and  $0.039$ , respectively) in milk from vegetarians (21; 18, 26 µg/l) and vegans (19; 18, 25 µg/l) compared with omnivores (17; 14, 20 µg/l). No statistically significant difference in Se between vegetarians and vegans was observed ( $P=0.42$ ). I was the next mineral approaching significant differences by maternal diet ( $P=0.10$ ).

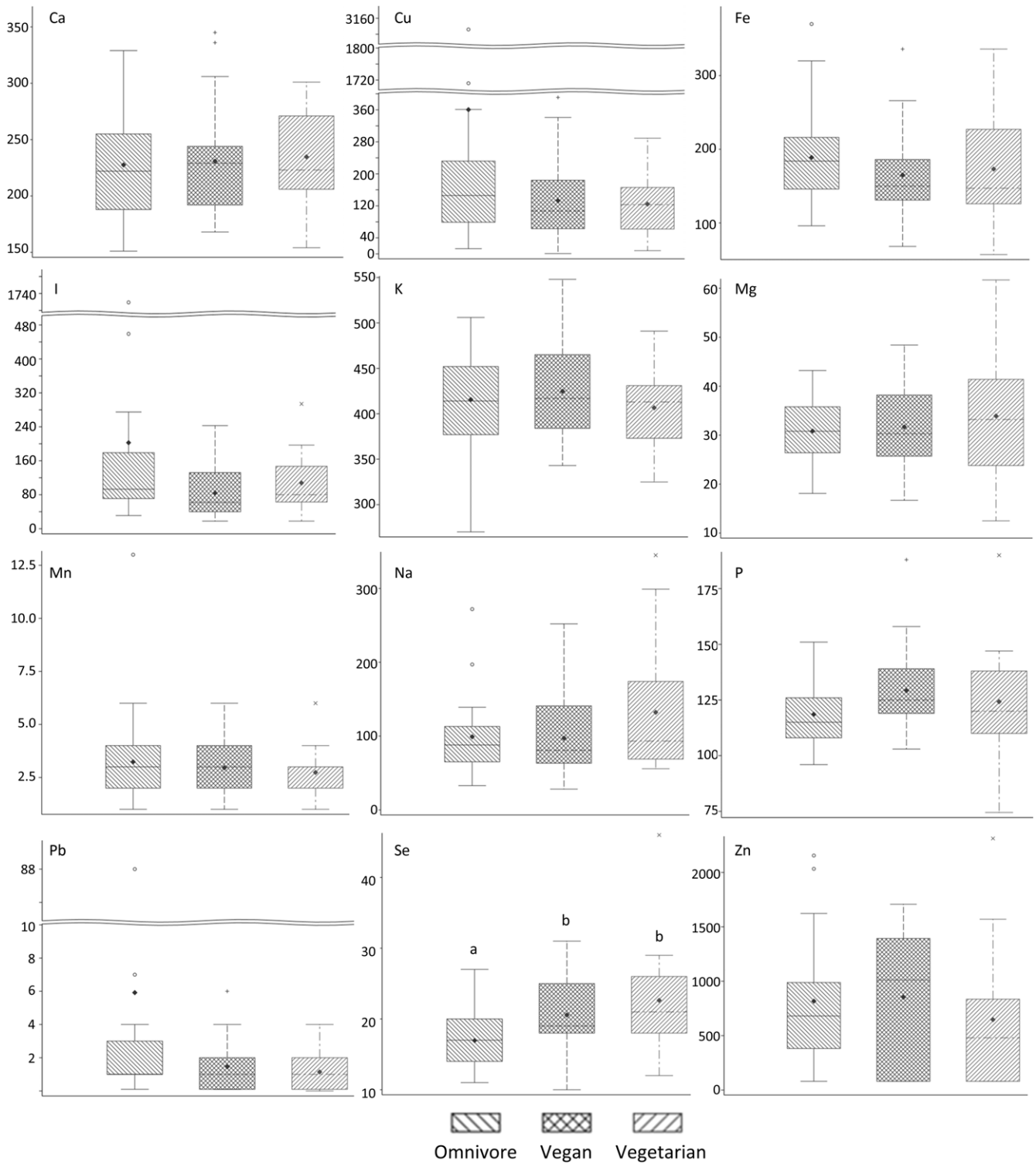
Variables used in the machine learning analyses included concentrations for twelve minerals, the use of multi-micronutrient or prenatal supplements (Y/N), maternal age, infant age

and maternal BMI. RFI was initially employed considering all sixteen variables and three diet categories, that is, vegan, vegetarian and omnivore. The accuracy, Cohen's kappa statistic and estimate of error rate for the RFI model were 40.1%, 0.10 and 63.5%, respectively. Considering the poor results, a new analysis was performed to include only the most relevant variables identified in the initial RFI model (i.e. se and BMI) and two classification groups: omnivore and plant-based diet. Accuracy, Cohen's kappa statistic and estimate of error rate significantly improved to 73.2%, 0.41 and 25.4%, respectively.

Because no single statistical method is capable of providing a perfect model for real samples, four additional techniques were applied to the data set. As shown in Fig. 2, the Boruta algorithm confirmed the RFI results by selecting se and BMI as the most significant variables for identifying a maternal diet as omnivore or plant-based. The ReliefF method again confirmed the RFI results, although it ranked BMI, Fe and se as the top three most important variables (Fig. 3). The top five variables ranked by the SVM-RFE method in decreasing order were se, Fe, P, I and Cu. Finally, the penalised logistic regression with Lasso penalty identified se, I and Fe as important variables, with coefficient values of  $-1.28$ ,  $1.13$  and  $0.93$ , respectively. With a model intercept value of  $-0.90$ , note that I and Fe are relatively higher (positive coefficients), and se is relatively lower (negative coefficients) for omnivores compared with individuals following a plant-based diet.

#### Discussion

We assessed breast milk mineral composition from US women following plant-based (vegan and vegetarian) and omnivore diets and found that Se was consistently identified as differing using a variety of statistical and machine learning techniques.

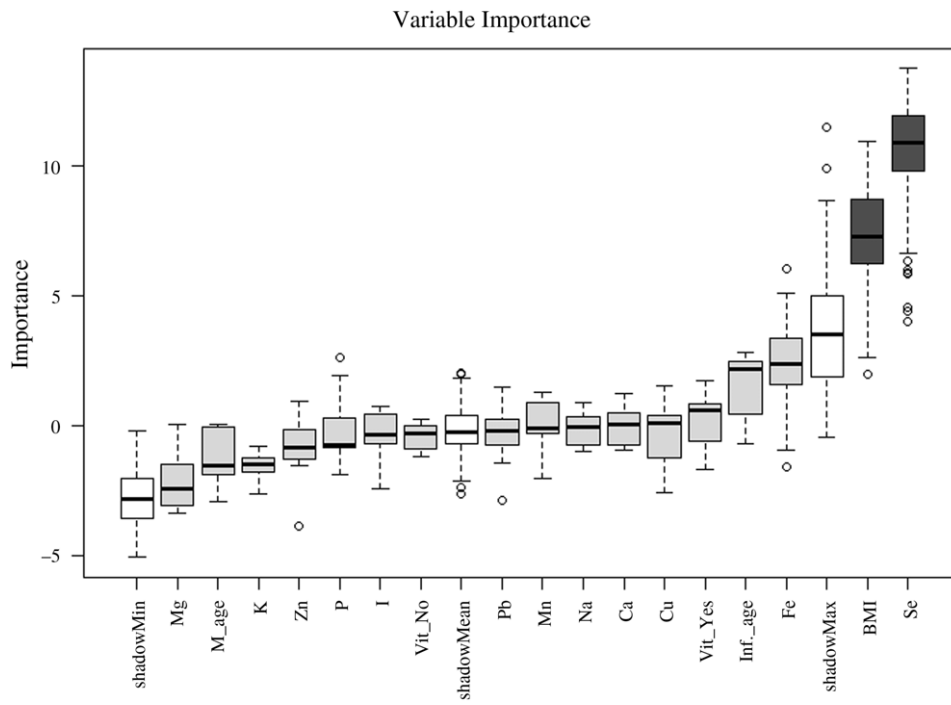


**Fig. 1.** Box and whisker plots showing the distribution of minerals in breast milk by maternal diet pattern. Plots with a different letter are statistically different based on a Kruskal–Wallis analysis ( $P = 0.007$ ) followed by a Dunn’s test for multiple comparisons (omnivore–vegan,  $P = 0.039$ ; omnivore–vegetarian,  $P = 0.008$  and vegan–vegetarian,  $P = 0.42$ ).

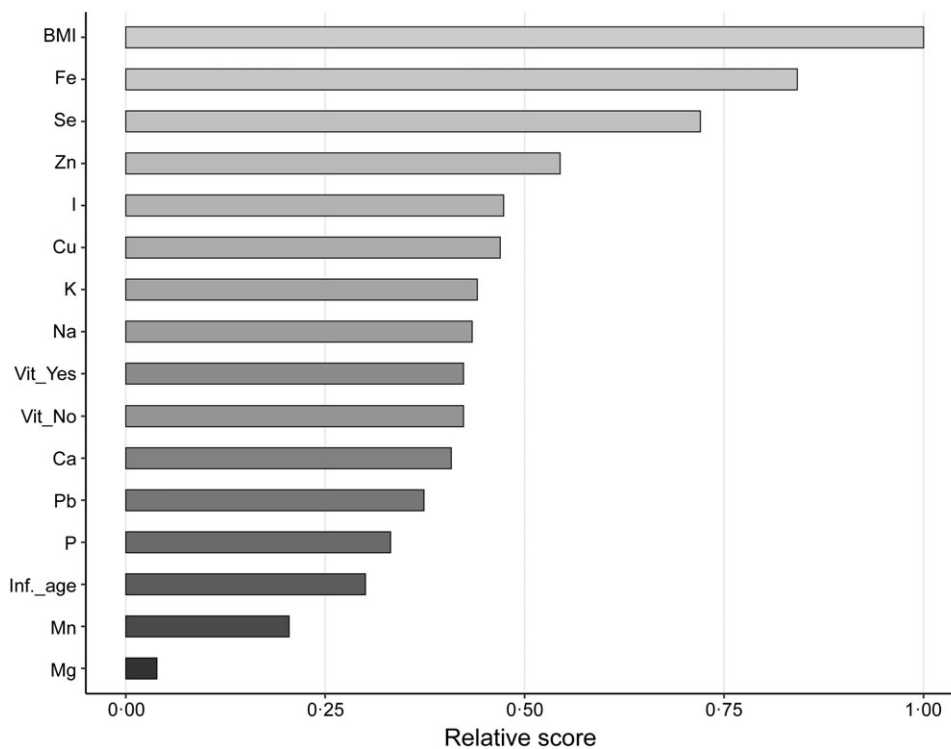
This approach strengthens our conclusion that Se may be higher in breast milk from women in the USA following a plant-based diet than those following an omnivore diet. Our finding that Se is higher in the milk of women following a plant-based diet agrees

with findings by Debski *et al.* in milk collected from lacto-ovo-vegetarians ( $n = 26$ ) and omnivores ( $n = 12$ )<sup>(11)</sup> but is surprising given reports of low Se intake in adults following a vegan diet<sup>(3,29)</sup>. It is possible that the high rate of supplement use in





**Fig. 2.** Feature ranking with Boruta (algorithm based on random forests) showing milk Se levels and mother's BMI are important to identify the different types of diet (omnivore or plant-based). Inf\_age, age of infant when sample collected; M\_age, maternal age; Vit\_No, did not use multi-micronutrient or prenatal vitamin; Vit\_Yes, did use a multi-micronutrient or prenatal vitamin.



**Fig. 3.** ReliefF algorithm showing BMI and Fe as the most important features to identify the different types of diet and their effect on milk elemental concentrations. Se and Zn present moderate importance for classifying the different diets (omnivore v. plant-based). Inf\_age, age of infant when sample collected; Vit\_No, did not use multi-micronutrient or prenatal vitamin; Vit\_Yes, did use a multi-micronutrient or prenatal vitamin.

our study population influenced the breast milk Se concentrations we observed, as Se from supplements has been shown to influence breast milk Se concentrations<sup>(30)</sup>. We did not collect detailed information about supplements used or their Se content. To better understand the Se status of lactating individuals following different diet patterns, future studies should also measure maternal serum Se concentrations.

There is limited information regarding composition of other minerals in breast milk based on maternal diet pattern. Finley *et al.* conducted a semi-longitudinal study of 222 samples of milk collected from fifty-two well-nourished participants between 1 and 20 months postpartum. Individuals were classified as vegetarian ( $n$  26) if they consumed no meat and only consumed fish no more than twice per month; semi-vegetarian ( $n$  6) if they consumed fish more frequently and non-vegetarian ( $n$  20) if they consumed meat<sup>(10)</sup>. When comparing milk composition between vegetarians and non-vegetarians, they reported no difference in Ca, Cu, Fe, Mg, K, Na and Zn. Using traditional statistical analysis, we also found no significant difference in these seven minerals by maternal diet pattern, along with no difference in I, Pb, Mn and P. However, using multiple machine learning techniques to address the limitations of a relatively small sample size, Se was repeatedly identified as a predictor of maternal diet-pattern, with Fe and I showing some marginal importance.

Our finding that maternal plant-based diet patterns may influence the Se composition of breast milk warrants further investigation as a potential factor for predicting infant nutritional exposure. Future research should distinguish maternal mineral intake from diet *v.* supplementation, and maternal mineral status, which we were unable to do in this study. A broader study with a larger number of samples may identify other minerals such as I as important breast milk biomarkers for maternal diet. While several of our machine learning models identified Fe as a potential predictor of maternal diet pattern, this is not supported in the literature and highlights the importance of assessing maternal intake and status of micronutrients in addition to milk composition<sup>(30,31)</sup>.

Longitudinal studies suggest that many minerals decline in breast milk for the first 1–2 months postpartum, including Ca, Cu, I, Fe, K, P and Zn<sup>(12,32)</sup>. In our cross-sectional study, the only early postpartum change we observed was a weak negative correlation between Zn and infant age. The youngest infant age in our study was 3.5 weeks postpartum, suggesting we may not have captured the early postpartum window where major time-related changes in mineral composition would be expected. Other minerals, such as Na and Se, have been reported to increase in late lactation stages, likely due to gradual weaning and involution of the mammary gland<sup>(12,32)</sup>. In a longitudinal cohort of breast-feeding women, significant increases in breast milk Na concentration were not observed until 15 months postpartum when daily breast-feeding sessions had dropped below five, suggesting that breast-feeding intensity more than lactation stage influences involution biomarkers<sup>(33)</sup>. We did not measure breast-feeding frequency, which would be valuable to consider in future studies to probe for a potential weaning-effect. It is unclear whether our findings of higher Se in milk from vegetarians compared with omnivores are primarily related to

differences in diet or lactation stage between the two groups. This would not likely explain differences in Se concentrations between vegans and omnivores, as lactation stage did not differ in our study between these groups.

Limitations to our study include the cross-sectional design, which did not capture some of the time-dependent changes in breast milk previously reported in the literature. Breast-feeding frequency and breast milk volume were not collected, which might have helped identify compositional changes associated with weaning<sup>(34)</sup>. We did not assess mineral intake from food or supplements. Our sample size ( $n$  63) was small and had limited racial diversity; however, it was larger than the two previous studies on minerals in the breast milk from vegan and vegetarian mothers. Additionally, the average breast-feeding duration of our study population was approximately 10 months postpartum, which is higher than US breast-feeding rates (55.8% of infants receive breast milk at 6 months postpartum)<sup>(35)</sup> and may limit generalisability.

This study used traditional statistics and machine learning techniques to minimise the sample size limitation (e.g. the RFI model was applied with a 5-fold cross validation and 10 repetitions) and consistently highlighted the importance of Se to differentiate breast milk composition by maternal plant-based diet pattern. It is important to emphasise that no single machine learning model will be perfect at finding all relevant variables, hence the need for applying several methods and using their results to reach better conclusions. In addition to Se, which has been identified as an important variable by all methods evaluated here, Fe was identified as important by three (ReliefF, SVM-RFE and penalised logistic regression with Lasso penalty) out of five machine learning methods, and I by two (SVM-RFE and penalised logistic regression with Lasso penalty) out of five. Therefore, even though no statistical difference has been found for Fe and I using descriptive statistics, the machine learning results suggest these elements should be further investigated. Future research on the impact of maternal plant-based diet on breast milk composition is warranted, using a larger sample size and a longitudinal design. Future studies should also consider maternal mineral intake, maternal mineral status and infant outcome measures including growth and development.

### Supplementary material

For supplementary material accompanying this paper visit <https://doi.org/10.1017/S0007114522004007>

### Acknowledgements

The authors are grateful to the individuals who participated in this study.

This research was supported by grants from the Vegetarian Nutrition Dietetic Practice Group, Academy of Nutrition and Dietetics Foundation to RP and MTP; and the National Science Foundation's Major Research Instrumentation Program (NSF MRI, grant CHE-1531698) to GLD.

R. P. and M. T. P. conceptualised and designed the study; R. P. and M. T. P. collected study samples; G. L. D. and N. J. analysed

study samples; G. L. D. and M. T. P. analysed the data; J. C., G. L. D., R. P. and M. T. P. interpreted the findings; J. C. and M. T. P. drafted the manuscript; all authors reviewed the manuscript.

The authors have no conflicts of interest to report.

## References

- Gallup (2018) Inc G. Snapshot: Few Americans Vegetarian or Vegan. <https://news.gallup.com/poll/238328/snapshot-few-americans-vegetarian-vegan.aspx> (accessed April 2022).
- Neufingerl N & Eilander A (2021) Nutrient intake and status in adults consuming plant-based diets compared to meat-eaters: a systematic review. *Nutrients* **14**, 29.
- Bakaloudi DR, Halloran A, Ripplin HL, *et al.* (2021) Intake and adequacy of the vegan diet. A systematic review of the evidence. *Clin Nutr* **40**, 3503–3521.
- Haider LM, Schwingshackl L, Hoffmann G, *et al.* (2018) The effect of vegetarian diets on iron status in adults: a systematic review and meta-analysis. *Crit Rev Food Sci Nutr* **58**, 1359–1374.
- Pawlak R, Berger J & Hines I (2016) Iron status of vegetarian adults: a review of literature. *Am J Lifestyle Med* **12**, 486–498.
- Foster M, Chu A, Petocz P, *et al.* (2013) Effect of vegetarian diets on zinc status: a systematic review and meta-analysis of studies in humans. *J Sci Food Agric* **93**, 2362–2371.
- Foster M, Herulah UN, Prasad A, *et al.* (2015) Zinc status of vegetarians during pregnancy: a systematic review of observational studies and meta-analysis of zinc intake. *Nutrients* **7**, 4512–4525.
- Raiten DJ, Raghavan R, Porter A, *et al.* (2014) Executive summary: evaluating the evidence base to support the inclusion of infants and children from birth to 24 month of age in the Dietary Guidelines for Americans – ‘the B-24 Projects.’ *Am J Clin Nutr* **99**, 663S–691S.
- Karcz K & Królak-Olejnik B (2021) Vegan or vegetarian diet and breast milk composition – a systematic review. *Crit Rev Food Sci Nutr* **61**, 1081–1098.
- Finley DA, Lönnnerdal B, Dewey KG, *et al.* (1985) Inorganic constituents of breast milk from vegetarian and nonvegetarian women: relationships with each other and with organic constituents. *J Nutr* **115**, 772–781.
- Debski B, Finley DA, Picciano MF, *et al.* (1989) Selenium content and glutathione peroxidase activity of milk from vegetarian and nonvegetarian women. *J Nutr* **119**, 215–220.
- Bilston-John SH, Narayanan A, Lai CT, *et al.* (2021) Macro- and trace-element intake from human milk in Australian infants: inadequacy with respect to national recommendations. *Nutrients* **13**, 3548.
- Pawlak R, Vos P, Shahab-ferdows S, *et al.* (2018) Vitamin B<sub>12</sub> content in breast milk of vegan, vegetarian and non-vegetarian lactating women in the United States. *Am J Clin Nutr* **108**, 525–531.
- Perrin MT, Pawlak R, Dean LL, *et al.* (2019) A cross-sectional study of fatty acids and Brain-Derived Neurotrophic Factor (BDNF) in human milk from lactating women following vegan, vegetarian, and omnivore diets. *Eur J Nutr* **26**, 2401–2410.
- Dubascoux S, Andrey D, Vigo M, *et al.* (2018) Validation of a dilute and shoot method for quantification of 12 elements by inductively coupled plasma tandem mass spectrometry in human milk and in cow milk preparations. *J Trace Elem Med Biol* **49**, 19–26.
- Donati G (2022) Advanced statistical tools and machine learning applied to elemental analysis associated with medical conditions. In *ICP-MS and Trace Element Analysis as Tools for Better Understanding Medical Conditions*, pp. 53–88 [M Zezzi-Arruda & JR Jesus, editors]. Amsterdam: Elsevier.
- Breiman L (2001) Random forests. *Mach Learn* **45**, 5–32.
- Goldstein BA, Polley EC & Briggs FBS (2011) Random forests for genetic association studies. *Stat Appl Genet Mol Biol* **10**, 32.
- Kursa MB & Rudnicki WR (2010) Feature selection with the Boruta package. *J Stat Softw* **16**, 1–13.
- O’Sullivan B & Woodrige M (editors) (2015–2020) Artificial Intelligence: Foundations, Theory, and Algorithms. <https://www.springer.com/series/13900> (accessed June 2022).
- Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In *Machine Learning: ECML-94*, pp. 171–182 [F Bergadano & L De Raedt, editors]. Berlin: Springer.
- Guyon I, Weston J, Barnhill S, *et al.* (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* **46**, 389–422.
- Guyon I & Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* **3**, 1157–1182.
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Stat Methodol* **58**, 267–288.
- Kassambara A (2018) *Machine Learning Essentials: Practical Guide in R*, 1st ed. Frankreich: CreateSpace Independent Publishing Platform.
- Succop PA, Clark S, Chen M, *et al.* (2004) Imputation of data values that are less than a detection limit. *J Occup Environ Hyg* **1**, 436–441.
- R-Project R: The R Project for Statistical Computing. <https://www.r-project.org/> (accessed June 2022).
- Schober P, Boer C & Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. *Anesth Analg* **126**, 1763–1768.
- Hess JM (2022) Modeling dairy-free vegetarian and vegan USDA food patterns for nonpregnant, nonlactating adults. *J Nutr* **152**, 2097–2108.
- Keikha M, Shayan-Moghadam R, Bahreynian M, *et al.* (2021) Nutritional supplements and mother’s milk composition: a systematic review of interventional studies. *Int Breastfeed J* **16**, 1–30.
- Domellöf M, Lönnnerdal B, Dewey KG, *et al.* (2004) Iron, zinc, and copper concentrations in breast milk are independent of maternal mineral status. *Am J Clin Nutr* **79**, 111–115.
- Allen JC, Keller RP, Archer P, *et al.* (1991) Studies in human lactation: milk composition and daily secretion rates of macronutrients in the first year of lactation. *Am J Clin Nutr* **54**, 69–80.
- Perrin MT, Fogleman AD, Newburg DS, *et al.* (2017) A longitudinal study of human milk composition in the second year postpartum: implications for human milk banking. *Matern Child Nutr* **13**, e12239.
- Bilston-John SH, Narayanan A, Tat Lai C, *et al.* (2021) Daily and within-feed variation of macro- and trace-element concentrations in human milk and implications for sampling. *Food Chem* **363**, 130179.
- Centers for Disease Control and Prevention (2022) Breastfeeding Report Card (Internet). Centers for Disease Control and Prevention. <https://www.cdc.gov/breastfeeding/data/reportcard.htm> (accessed November 2022).