Personal Diet Inference through Mixture of Concentration Constraints and Intake Preferences

Jari Turkia[*,†], Ursula Schwab[‡,§], Ville Hautamäki[*]

Abstract. Maintaining proper nutrition is crucial for preserving health and preventing diseases. However, what constitutes proper nutrition may vary among individuals; recent evidence indicates that the effects of diet and also nutrients can differ considerably, driven by unique personal characteristics. This variability is reflected in blood markers, such as cholesterol and insulin concentrations, which respond differently to nutrient intake. This study captures this variation through a hierarchical, multivariate model that relates nutrient intake to various concentrations and applies it to personal diet inference.

Our main contribution is a method for personal diet inference using a conditional two-component Bayesian mixture model that leverages the Nordic Nutrition Recommendations 2023 as a prior for healthy intake and infers personalized recommendations as posterior distributions. The first component identifies dietary options predicted to produce healthy levels across all considered blood markers, while the second component determines the diet closest to predefined personal preferences within these valid options. This preference component minimizes required dietary adjustments to promote patient adherence while providing well-defined values for nutrients less critical to blood marker regulation.

We evaluated our method using nutritional data from two different studies: one in prediabetic patients and the other in patients with kidney dysfunction. Through numerical simulations, we demonstrated that personalized diets derived from our model could restore normal plasma concentrations. As the results align with the current nutritional literature, the use of Bayesian inference could open new opportunities to advance research using observational nutrition data. We aim for future clinical studies to validate these computational insights for statistically justified and more precise personal nutritional counseling.

Keywords: Bayesian statistics, Computational systems biology, Diet, Hierarchical models, Multivariate models, Mixture models, Nutrition, Personalization, Human.

[*]School of Computing, University of Eastern Finland, Joensuu, Finland

[†]CGI Suomi Oy, Joensuu, Finland, jari.turkia@cgi.com

[‡]School of Medicine, Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

[§]Wellbeing Services County of North Savo, Kuopio University Hospital, Department of Medicine, Endocrinology and Clinical Nutrition, Kuopio, Finland

1 Introduction

Proper nutrition plays a vital role in maintaining health and preventing disease (World Health Organization (2024)). However, it has become increasingly evident that the impact of nutrition varies between individuals due to differences in nutrient metabolism and regulation (Cecil and Barton (2020), Boorsma et al. (2017)). Personalized nutrition (Bush et al. (2020)) aims to address this variability by providing personalized dietary guidance tailored to individual characteristics, going beyond generic recommendations (Boorsma et al. (2017)).

Although nutritionists provide personal guidance, accurately determining individual responses and creating evidence-based diets remains challenging without rigorous statistical tools. Personalized guidance is particularly crucial for individuals with chronic conditions like kidney dysfunction. For example, the National Kidney Foundation's KDOQI guidelines recommend personalized dietary adjustments to maintain balanced serum concentrations, but leave implementation to nutritionists (Ikizler et al. (2020)). Statistical models of patient responses could greatly improve the development of evidence-based dietary recommendations.

Predictive modeling, like decision tree-based methods, has shown promise in personalized nutrition. These models effectively manage postprandial glucose responses in prediabetes (Zeevi et al. (2015), Ben-Yacov et al. (2021)) and newly diagnosed type 2 diabetes mellitus (Rein et al. (2022)), outperforming generic healthy diets. Personalized guidance also promotes better food choices compared to generic dietary guidelines (Celis-Morales et al. (2016)), with greater benefits observed in individuals motivated to improve their health (Livingstone et al. (2020)).

A systematic review of statistical methods in personalized nutrition (Kirk et al. (2021)) highlights algorithms, including decision tree-based methods, that perform well in recommending whole meals or dietary patterns. However, these approaches often overlook the critical challenge of balancing plasma concentrations, an actionable measure of personal health (Martini et al. (2017)). Addressing this gap requires methods that model nutrient-concentration relationships rather than simply recommending diets. Gleason and Harris (Gleason and Harris (2019)) identified the Bayesian framework as ideal for personalized nutrition due to its ability to integrate prior knowledge, quantify uncertainty, and make probabilistic inferences. By capturing the relationships between nutrient intake and plasma concentrations within a probabilistic framework, Bayesian methods can provide personalized dietary recommendations that not only meet intake preferences, but also ensure optimal plasma concentration levels. Despite these advantages, Bayesian methods in this field remain largely untapped, offering significant opportunities for further research and development, which this study aims to address.

Our previous work demonstrated that Bayesian hierarchical models can effectively capture individual nutrient effects (Turkia et al. (2021)). Building on the biology of personal nutrition (Boorsma et al. (2017)), we developed a model that uses nutrient intake as a predictor of plasma concentrations, with latent variables representing individual differences in how nutrients are absorbed and metabolized. Using this

framework, we proposed a Bayesian diet inference method that recommends optimal intake levels for selected nutrients while maintaining the remainder of the current diet (Turkia et al. (2024)). Simulations showed that this approach could effectively manage plasma concentrations of phosphate and potassium, preventing hyperphosphatemia and hyperkalemia in renal patients. However, these targeted adjustments proved insufficient to achieve the desired concentration levels for all patients, underscoring the need for a more comprehensive approach and motivating the present study.

In this study, we generalize our previous results to address concentration control by inferring the most effective personal intake adjustments. The main contribution of this work is a Bayesian inference model that personalizes diet recommendations through posterior inference, using the Nordic Nutrition Recommendation 2023 (NNR 2023) (Blomhoff et al. (2023)) as priors for healthy intake. Our model identifies optimal intake levels that satisfy specified concentration constraints while aligning with individual dietary preferences. This is achieved through a conditional mixture of two components: the first prioritizes diets that meet the concentration constraints, and the second selects the configuration that best matches personal preferences within these options.

The concentration constraints are implemented using smooth sigmoid functions, enabling the model to tolerate invalid prior values and support gradient-based estimation methods. Sensitivity analysis showed that such invalid values could occur at the edges of intake ranges allowed by general dietary recommendations (e.g., NNR 2023), where personal nutrient effects predict concentrations outside the recommended limits. These predictions caused the inference model to assign zero probability to such diets, highlighting not only the computational challenge but also the limitations of broad dietary guidelines in addressing individual variability. By smoothing the concentration constraints, the model can initialize from partially invalid intake values and iteratively adjust toward valid diets. The preference component complements this by addressing identifiability issues: while multiple intake combinations may yield healthy concentrations, incorporating a clear preference establishes a well-defined mode in the posterior distribution. In this work, the personal preference is set for the current diet, assuming that minimal but effective modifications improve the practicality of personalized nutrition. The resulting mixture model ensures that the expected values of intake distributions align with the preference, while the concentration constraints guide the posterior towards healthy concentrations, maintaining the entire 90%-credible interval within acceptable ranges.

We aim to demonstrate how personalized diets, guided by established dietary guidelines, can theoretically balance plasma concentrations in individuals with unhealthy levels. To evaluate this, we apply our inference model to two patient groups: prediabetes and kidney dysfunction, utilizing observational data from dietary records and corresponding laboratory analyses. The simulation results show that our model generates personalized diet recommendations that balance key plasma concentrations with minimal dietary adjustments, supporting feasibility and adherence.

We adhere to the Bayesian Analysis Reporting Guidelines (BARG, Kruschke (2021)), providing comprehensive details on the data, code, models, computational

processes, posterior distributions, inferences, and sensitivity analyses for transparency and reproducibility.

In summary, the key contributions of this work are:

- Demonstrating that a hierarchical multivariate model effectively captures individual variations in nutrient effects at multiple plasma concentrations.
- Developing a conditional two-component mixture model that integrates general nutritional guidelines with individual preferences to infer personalized dietary recommendations as posterior distributions.
- Validating the proposed method through simulations on two patient groups, showing that in the simulation setting it can balance plasma concentrations within healthy ranges by leveraging the estimated functional roles of nutrients.

2 The inference process and applied nutritional data

Our process for personal diet inference involves two main steps: constructing personal models of concentration compositions and inferring nutrient intake likely to modify the concentrations as required. We posit that the primary source of the concentration variance is the personal effects of the nutrient intake observed in our previous research (Turkia et al. (2021, 2024)). Our method requires repeated personal food records and laboratory analyses of concentrations from the same period within a few days to estimate these effects. In addition to personally repeated observations, these data can include nested hierarchies, such as different treatments or relevant groupings. In addition to nutritional effects, we estimate current intake levels from observational data, which we later use to define personal preferences. In the inference phase, we do not use the observational data directly. Instead, we assign prior proposal distributions to the inferred nutrients based on general nutritional recommendations and derive their personalized posterior distributions. Figure 1 provides an overview of this process and the data involved.

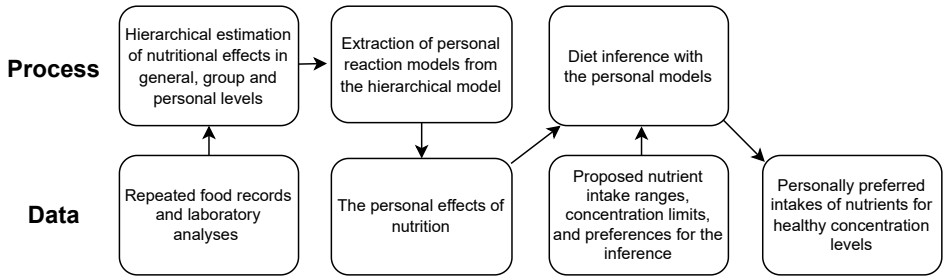

Figure 1: The process of inferring personal diet and the data involved in each step.

To evaluate the effectiveness of our method, we analyze two datasets that align with our described structure but involve different patients and variables. The first dataset,

Sysdimet, originates from a dietary intervention study of the same name (Lankinen et al. (2011)). This study included 106 subjects with impaired glucose metabolism and features of the metabolic syndrome. Data collection occurred at four time points over 12 weeks, during which subjects kept food records and provided personal information and laboratory results. The primary goal of personalized nutrition in this context was to achieve reference values for specific fasting plasma concentrations, including high-density lipoprotein (HDL-C), low-density lipoprotein (LDL-C), total cholesterol (TC), insulin, and glucose. Predictors of concentration compositions encompass sex, cholesterol-lowering medication status, and 20 different nutrients, all detailed in Table 1. While the original Sysdimet trial randomly assigned subjects to groups to study the effects of different diets, our analysis treats subjects as independent, disregarding the randomization to groups.

The second dataset (Dialysis) comprises data from 37 end-stage renal failure patients undergoing dialysis (Turkia et al. (2024)). The cohort includes 15 women and 22 men, aged 26 to 81 years (mean age: 61). For each patient, two observations were collected three months apart. Patients were interviewed about their dietary intake over the preceding 48 hours during each observation period, and laboratory tests were conducted within a week of the interview. For this analysis, we focus on three target concentrations from the laboratory tests: plasma potassium, fasting plasma phosphorous, and plasma albumin. Our predictors for these concentrations include dietary nutrients and several patient characteristics: sex, type of dialysis treatment, and current medications (including blood lipid medication, phosphate binder medication, diabetes medication, and hydroxycholecalciferol supplementation). Table 1 presents the nutrients considered in both datasets, along with their mean intake and standard deviation.

The inference phase involves establishing healthy nutrient ranges with adjusted intake levels, as detailed in Table 1 for both datasets. These ranges are derived from the Nordic Nutrient Recommendations 2023 (Blomhoff et al. (2023)), which provide key reference values for each nutrient: recommended intake (RI) or adequate intake (AI), average requirement (AR) or provisional AR, and upper limit (UL). We aligned our ranges to the most suitable lower bound depending on the nutrient. The upper bounds typically align with the UL, representing the maximum intake generally tolerated without adverse effects. While our primary goal is to determine personally optimal intake levels, these established ranges serve as crucial guidelines to ensure our recommendations remain within the relevant range. In addition to these limits, we also incorporate general intake recommendations. The inference model can be configured to favor these recommended levels, helping prevent extreme values in the final personalized recommendations.

Table 1: Nutrients analyzed and included in the inference of personal recommendations in one or both studied examples. The prior distributions for the nutrient random variables are implemented as a mixture of two distributions to balance between fully personalized recommendations and general guidelines. The proposed range, from Q_r^l to Q_r^u, defines a uniform distribution, while a normal distribution centered at a specified general recommendation (Q_r^p) can be enforced with a hyperparameter λ_6 to influence the degree of adherence to general recommendations and steer personal recommendations away from extreme values when necessary.

	General		Study mean ± SD	
Nutrient	**Proposed range**	**Recommendation**	**Sysdimet**	**Dialysis**
Sucrose (E%)	0 - 10	8	6.0 ± 2.6	9.2 ± 4.4
Protein (E%)	10 - 20	15	18.5 ± 3.0	15.1 ± 3.0
Carbohydrates (E%)	40 - 60	50	47.4 ± 6.1	43.6 ± 7.7
Saturated fatty acids (E%)	7 - 10	8.5	11.8 ± 2.9	13.7 ± 4.1
Monounsaturated fatty acids (E%)	10 - 20	15	9.7 ± 2.7	14.7 ± 3.6
Polyunsaturated fatty acids (E%)	5 - 10	7	4.7 ± 1.5	7.1 ± 2.4
Fiber (g/d)	20 - 55	30	19.0 ± 5.6	16.7 ± 6.2
Vitamin D (ug/d)	2.5 - 100	10	6.7 ± 3.5	7.6 ± 5.0
Calcium (mg/d)	400 - 2500	950	-	570 ± 301
Potassium (mg/d)	1600 - 3700	3500	-	2785 ± 961
Salt (mg/d)	1500 - 5750	3750	-	6560 ± 3099
Phosphorous (mg/d)	300 - 3000	520	-	1042 ± 379
Water (ml/d)	500 - 2500	1500	-	1804 ± 607
Linoleic acid (E%)	2.5 - 9		3.3 ± 1.1	-
Alpha-linolenic acid (E%)	0.5 - 5		0.7 ± 0.3	-
Cholesterol (mg/d)	0 - 300		247 ± 99	-
Folate (ug/d)	100 - 1000	330	229 ± 74	-
Vitamin C (mg/d)	10 - 1000	100	101 ± 66	-
EPA (mg/d)	0 - 2.5		0.1 ± 0.1	-
DHA (mg/d)	0 - 2.5		0.2 ± 0.2	-

EPA = Eicosapentaenoic acid, DHA = Docosahexaenoic acid

Table 2 shows the normal ranges for the concentrations targeted in both studies. The baseline for these ranges follows Finnish national recommendations (Duodecim (2024)). However, the recommended upper limits of insulin (20.0 mU/l), LDL-cholesterol (3.0 mmol/l), and total cholesterol (5.0 mmol/l) have been slightly relaxed as several individuals in the Sysdimet cohort could not reach these thresholds in our simulated dietary adjustments. Additionally, fasting plasma glucose levels in this cohort tend to cluster near the upper boundary of the normal range, making the primary objective of our recommendation model to reduce glucose levels while maintaining other concentrations within their respective normal ranges. In contrast, the dialysis cohort presents a different challenge, with average albumin levels frequently falling below the normal range. Here, our model prioritizes increasing albumin levels while carefully managing potassium and phosphate concentrations to ensure overall balance.

Table 2: Targeted normal ranges for plasma concentrations may vary by age and sex. The problematic concentration limits for each study are highlighted in bold. Normal ranges follow Finnish national recommendations (Duodecim (2024)), except that the upper limits of insulin, LDL- and total cholesterol have been slightly relaxed in the Sysdimet cohort to allow for more feasible dietary adjustments.

			Study mean ± SD	
Plasma concentration	**Target range**	**Target group**	**Sysdimet**	**Dialysis**
HDL cholesterol	0.8 - 2.1 mmol/l	Men	1.5 ± 0.3	-
HDL cholesterol	1.0 - 2.7 mmol/l	Women	1.4 ± 0.4	-
LDL cholesterol	2.0 - **3.5** mmol/l	Everyone	3.2 ± 0.8	-
Total cholesterol	3.9 - **5.5** mmol/l	Everyone	5.2 ± 0.9	-
Insulin	2.6 - 25 mU/l	Everyone	12.1 ± 6.3	-
Glucose	4 - **6** mmol/l	Everyone	**6.0** ± 0.5	-
Potassium	3.4 - 4.7 mmol/l	Dialysis patients	-	4.4 ± 0.6
Phosphate	1.13 - 1.78 mmol/l	Dialysis patients	-	1.66 ± 0.48
Albumin	**36** - 48 g/l	39 years or younger	-	**37** ± 0
Albumin	**36** - 45 g/l	40-69 years old	-	**34** ± 5.3
Albumin	**34** - 45 g/l	70 years and older	-	**32** ± 5.3

We assume that personal factors and multiple nutrients in current diets influence plasma concentrations. Although factors such as treatments, age, and sex remain fixed, nutrient intakes are adjustable within recommended ranges. These ranges, based on established guidelines, ensure that personalized recommendations address specific concentration challenges while staying within healthy limits.

3 Personal Diet Inference

The diet inference method aims to identify diet compositions that maintain all plasma concentrations within their normal ranges, with a given probability level denoted c. For an individual k, this inference task is illustrated as a directed graphical model G_k in Fig. 1, following the notation of Koller and Friedman (2009). The process consists of two main steps. First, personal relationships between nutrient intake and the corresponding plasma concentrations are estimated using a multivariate hierarchical Bayesian regression. Subsequently, the diet inference is performed within the estimated graph $\hat{G}_k$ imposing lower and upper limits, Y_l^m and Y_u^m, on all $m = 1, \ldots, M$ concentrations, and the optimal intake of nutrients $\hat{Q}_r$ is inferred.

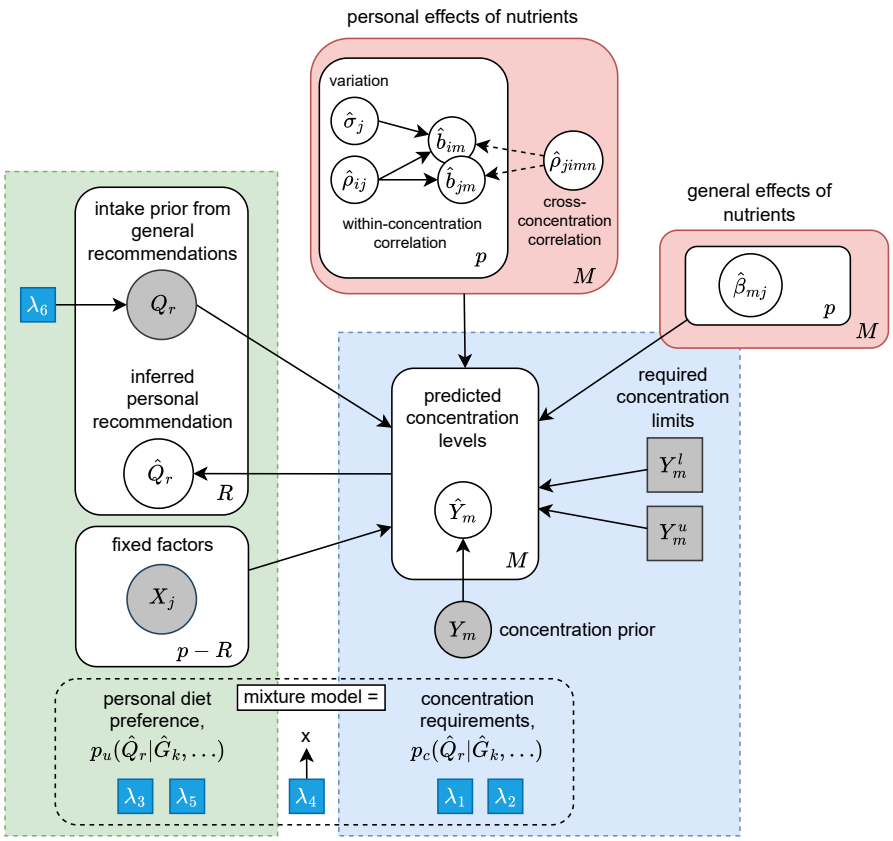

Figure 2: Graphical model $\hat{G}_k$ representing the personalized nutrition status and inference for individual k. The model encapsulates the prior belief about the relationship between nutrient intake and resulting concentrations (pink boxes). The personal nutrition recommendations $\hat{Q}_r$ are the posterior inference results of a mixture model, balancing two components: the concentration requirements (dashed blue box) and the personal intake preferences (dashed green box). Circular nodes represent random variables, while squared nodes denote fixed hyperparameters guiding the recommendations. Solid boxes with repetition indices in the bottom-right corner indicate plate notation for repeated variables.

3.1 Hierarchical Model of Concentrations

Concentrations are characterized by more individuals having higher concentration values than lower ones, resulting in a right-skewed distribution. Together with exclusively positive values, we assume that the Gamma distribution best captures this pattern

(Limpert and Stahel (2011)). Specifically, the i-th observation of concentration m for individual k is modeled as

$$Y_{kmi} \sim \text{Gamma}\left(\alpha_m, \frac{\alpha_m}{\mu_{kmi}}\right), \quad i = 1, \ldots, n_k; \quad k = 1, \ldots, K; \quad m = 1, \ldots, M, \tag{1}$$

where the concentration shape parameters α_m are estimated alongside the nutritional effects, and the rate parameters are given by α_m/μ_{kmi}, ensuring that mean of the distribution is μ_{kmi}. These expected concentration values are obtained using a hierarchical linear predictor

$$\mu_{kmi} = \beta_{m0} + b_{km0} + \sum_{j=1}^{p-R} X_{kji}\left(\beta_{mj} + b_{kmj}\right) + \sum_{r=1}^{R} Q_{kri}\left(\beta_{mr} + b_{kmr}\right), \tag{2}$$

where we intentionally separate the fixed predictors X_{kji} (e.g., medication, sex) from the inferred intakes Q_{kri} at observation i. The terms β_{m0} and b_{km0} represent the population-level and personal intercepts for concentration m, respectively. Similarly, the fixed predictors and inferred nutrient intakes have associated terms for the population-level effects β and personal deviations b from the population mean. To enhance interpretability, we use an identity link in the Gamma distribution's linear predictor, ensuring an additive relationship between predictors and concentrations. Additionally, all predictors are standardized before estimation, with both additivity and standardization facilitating consistent comparison across variables.

We model the concentrations as a system of seemingly unrelated regressions (SUR) (Mehtatalo and Lappi (2020)). Although there are no direct relationships between the dependent variables in the SUR models, the hierarchical effects introduce dependencies, allowing shared variability across the concentrations to be captured. This shared variability is realized by defining that the personal effects of each person k follow a common multivariate normal distribution

$$\mathbf{b}_k \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}\right), \tag{3}$$

where $\mathbf{\Sigma}$ is a square covariance matrix of dimension $M(p+1)$, with M denoting the number of concentrations and $p+1$ accounting for predictors, including personal intercepts. The covariance matrix is parameterized using a diagonal matrix of standard deviations $\mathbf{T}$ and a correlation matrix $\mathbf{C}$ with

$$\mathbf{\Sigma} = \mathbf{TCT}, \tag{4}$$

having $\mathbf{T} = \text{diag}\left(\sigma_1, \sigma_2, \ldots, \sigma_{M(p+1)}\right)$. The correlation matrix $\mathbf{C}$ captures the correlations between personal effects within and between concentrations. The diagonal blocks of $\mathbf{C}$ represent the correlations of personal effects within the same concentration m, while the off-diagonal blocks represent the correlations between different concentrations m and n (Turkia et al. (2024)).

Additional terms g_0 and g can be incorporated into the linear predictor to introduce a new nested level of effects, such as group-level effects. The effects at the group level $\mathbf{g}$

follow a similar multivariate distribution as the personal effects, with a covariance matrix defined and estimated separately $\mathbf{\Sigma}_g$. This extension introduces a distinct variance-covariance structure at the group level, allowing the model to capture shared variability between concentrations and different groups.

By distinguishing between fixed predictors and inferred nutrient intakes, we explicitly state that the intercepts and effects of the fixed predictors can be evaluated before diet inference, serving as concentration baselines. In contrast, the term for the queried nutrients, $Q_{kri}\,(\beta_{mr} + b_{kmr})$, can be seen as *contributions* of nutrients that the inference balances between concentrations.

3.2 Conditional Recommendation Distribution of Nutrients

The posterior distribution of inferred nutrient intakes Q_{kr} for person k, given the estimated personal graph $\hat{G}_k$ and the imposed concentration constraints, is defined

$$p(Q_{kr}|\hat{G}_k,\ P(Y_m^l < Y_m < Y_m^u) > c\,),\ r = 1, \ldots, R;\ m = 1, \ldots, M, \tag{5}$$

where the lower and upper limits, Y_m^l and Y_m^u, for concentration Y_m may also vary according to age and sex, as exemplified in Table 2. The prior distributions for the random variables of the nutrients are derived from the healthy intake limits specified in general nutritional guidelines. To balance between fully personalized recommendations and general guidelines, the prior is implemented with an adjustable mixture distribution

$$p(Q_r) \sim \frac{1}{2} \cdot \text{Uniform}(Q_r|Q_r^l, Q_r^u) + \frac{1}{2} \cdot \mathcal{N}(Q_r|Q_r^p, \sigma_r). \tag{6}$$

In this equation, Q_r^l and Q_r^u represent the normalized limits of a uniform distribution for healthy intake, mixed with a normal distribution for general intake recommendation. The mean of this normal distribution is set to Q_r^p, representing the single most recommended intake.

The personal effects of nutrients were estimated using a normalized intake scale, which also requires normalizing the proposed nutrient ranges by subtracting the mean and dividing the limits by the standard deviation. After inference, the original intake scales are restored in the nutrient recommendations.

There are several hyperparameters λ to adjust the posterior inference distribution, indicated with blue squares in Fig. 2. The hyperparameter λ_6 is used to scale the standard deviation of this distribution normal regardless of the length of the range of healthy intake of nutrients: $\sigma_r = (Q_r^u - Q_r^l)/\lambda_6$. This formulation allows for flexible adjustment of the influence of general recommendations in all nutrients while maintaining the specific ranges of healthy intake for each nutrient.

Implementing Confident Concentration Limits

Conditional concentration limits, $P(Y_m^l < Y_m < Y_m^u) > c$, can be simplified by acknowledging that predictive uncertainties, given by the shapes of the concentration

distributions, are already estimated in the personal model $\hat{G}_k$. This allows for the inference of expected values that maintain the required cumulative probability within the original limits.

For $Y_{kmi}|\hat{\alpha}_m, \hat{\mu}_{kmi} \sim \text{Gamma}\left(\hat{\alpha}_m, \frac{\hat{\alpha}_m}{\hat{\mu}_{kmi}}\right)$, the expected value of the distribution satisfies $\mathbb{E}[Y_{kmi}] = \hat{\mu}_{kmi}$. Here, $\hat{\alpha}_m$ is the estimated shape parameter, and $\hat{\mu}_{kmi}$ is both the expected value and the mean of the Gamma distribution, ensuring that $\hat{\mu}_{kmi}$ directly represents the average concentration for individual k, concentration m, and observation i.

To adjust the limits of the expected values, we calculate the quantiles of the Gamma distribution using its cumulative density function (CDF):

$$F_Y(y; \alpha, \beta) = \frac{\gamma(\alpha, \beta y)}{\Gamma(\alpha)}, \tag{7}$$

where $\Gamma(\alpha)$ is the gamma function of α, and $\gamma(\alpha, \beta y)$ is the lower incomplete gamma function. Using this CDF, we compute specific quantiles corresponding to the desired confidence level c and adjust the concentration limits to ensure compliance with the original constraints.

The adjusted limits for the expected values are derived as:

$$\begin{aligned} \mu_m^l &= F_Y^{-1}(1-c; \hat{\alpha}_m, \hat{\beta}_m), \\ \mu_m^u &= F_Y^{-1}(c; \hat{\alpha}_m, \hat{\beta}_m), \quad m = 1, \ldots, M, \end{aligned} \tag{8}$$

where F_Y^{-1} is the quantile function (inverse CDF) of the Gamma distribution, and $\hat{\beta}_m = \hat{\alpha}_m / \hat{\mu}_{kmi}$ is the estimated rate parameter. These adjusted limits constrain the expected concentration values $\hat{\mu}_{kmi}$ rather than the raw observations Y_{kmi}.

The transformation simplifies the joint conditional distribution to

$$p(Q_{kr}|\hat{G}_k, \mu_m^l < \hat{\mu}_m < \mu_m^u), \quad r = 1, \ldots, R, \quad m = 1, \ldots, M, \tag{9}$$

where $\hat{\mu}_m$ represents the expected concentration value of Y_m, and the constraints $\mu_m^l < \hat{\mu}_m < \mu_m^u$ ensure the desired cumulative probability. This approach effectively balances predictive uncertainty and concentration constraints by using only expected values $\hat{\mu}_{kmi}$.

Making the Limits Differentiable and Countable

The implementation of strict concentration limits using hard constraints as in Eq. (9) can pose challenges for optimization algorithms, particularly those relying on gradient information (Gelman et al., 2013). Furthermore, achieving the targeted concentration limits may not always be feasible with the initial proposal distributions of Q_r, as their priors from general recommendations can be broad and may result in concentrations outside the limits for some individuals. To address these problems, we implement the concentration limits with an inverse logit function that is differentiable and can be

adjusted to favor normal concentration ranges while maintaining support outside these limits:

$$\frac{1}{1 + \exp\left(\left(\mu_m^l - \hat{\mu}_m\right)\lambda_1\right)} \tag{10}$$

This function smoothly transitions from 0 to 1 as the expected concentration value $\hat{\mu}_m$ approaches the limit μ_m^l. The hyperparameter λ_1 controls the steepness of this transition, with higher values approximating a hard limit more closely.

Figure 3 illustrates this concept. The blue curve represents the inverse logit function rising from 0 to 1 as $\hat{\mu}_m$ approaches the adjusted lower limit μ_m^l, while the red curve shows a similar transition for the upper limit μ_m^u.

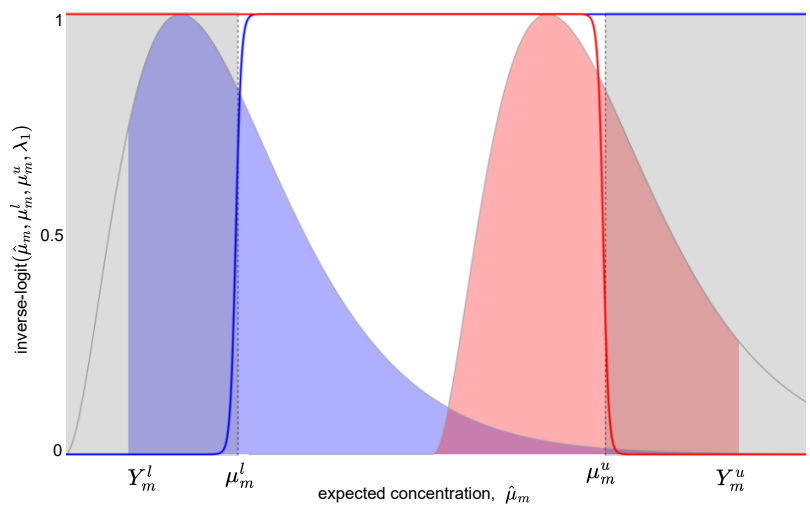

Figure 3: Illustration of transforming the original concentration limits Y_m^l and Y_m^u into corresponding expected values μ_m^l and μ_m^u. The inverse logit functions (blue and red curves) provide smooth transitions at the limit values, facilitating differentiable optimization.

To quantify the degree to which both adjusted limits, μ_m^l and μ_m^u, are met, we define a function that increases smoothly as the targeted limits are reached:

$$b(\hat{\mu}_m, \mu_m^l, \mu_m^u, \lambda_1) = \frac{1}{1 + \exp\left(\left(\mu_m^l - \hat{\mu}_m\right)\lambda_1\right)} + \frac{1}{1 + \exp\left(\left(\hat{\mu}_m - \mu_m^u\right)\lambda_1\right)} \tag{11}$$

that attains its maximum value of 2 when both limits are satisfied. The hyperparameter λ_1 determines the steepness of the transitions and should be calibrated to ensure non-zero probability for initial proposal values.

We then define a probability distribution associated with satisfying both limits of all the M concentrations as

$$p(Q_{kr}|\hat{G}_k,\ \mu_m^l < \hat{\mu}_m < \mu_m^u,\ \lambda_1,\ \lambda_2) = \mathcal{N}\left(\sum_{m=1}^{M} b(\hat{\mu}_m, \mu_m^l, \mu_m^u, \lambda_1) \,\middle|\, 2M, \lambda_2\right), \tag{12}$$

where the normal distribution with mean $2M$ reaches its maximum when all the limits are satisfied, and the hyperparameter λ_2 as controls the allowed standard deviation from these limits.

Although optimizing this distribution efficiently takes nutrient concentrations within the desired ranges, it can result in multiple equally valid diets. To resolve this ambiguity, we introduce a personal preference component to guide the selection among these possibilities.

3.3 Mixture of Concentration Control and Personal Preference

Although many diets can achieve normal nutrient concentrations, this abundance can make the inference distribution unidentifiable and complicate estimation. To address this, we incorporate a preference distribution into the joint recommendation, creating a mixture distribution with a single, most preferred mode. Importantly, this preference is applied only after all targeted concentration limits are met.

We denote the distribution of concentration constraints in Eq. (12) as mixture component p_{c} and the full conditional mixture model with

$$p(Q_{kr} \mid \hat{G}_k) = p(Q_r)\left[(1-w)\, p_{\mathrm{c}}(Q_{kr}|\hat{G}_k, \dots) \ + w\, p_{\mathrm{u}}(Q_{kr}|\hat{G}_k, u(\cdot))\right] \tag{13}$$

where w smoothly transitions the weight from the concentration constraint component p_{c} to the preference component p_{u} once all targeted concentration limits are met.

The preference function $u(\cdot)$ in p_{u} is only required to reach a unique maximum when the most preferred recommendation is achieved. This ensures the posterior distribution has a single, well-defined mode corresponding to the optimal intake under the given preference. The prior intake distributions $p(Q_r)$, derived from general dietary recommendations, are defined in Eq. (6).

The coefficient w is defined using the inverse logit function similar to concentration limits

$$w = \frac{1}{1 + \exp\left(\left((2M - 0.1) - b\left(\mu, \lambda_1\right)\right)\lambda_4\right)} \tag{14}$$

where $b(\mu, \lambda_1)$ is the count of satisfied limits from Eq. (11) and $2M$ is the total number of limits that it approaches. Subtracting 0.1 from this expected maximum is an implementation detail ensuring that w reaches 1 promptly when all limits are satisfied. Hyperparameter λ_4 adjusts the steepness of this conditional transition. This formulation provides a steep yet smooth transition between the mixture components, assigning weight to the preference component only after all concentration limits have been met.

Personal Preference of Minimal Diet Modification

This study defines personal preference as minimizing changes from the current diet. This objective is quantified using a preference function

$$u(\hat{G}_k, \lambda_5) = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{1}{R} \sum_{r=1}^{R} \frac{|X_{kr} - Q_{kr}|}{|Q_{kr}|} \cdot \frac{1}{1 + |\hat{\beta}_{mr}|} \right)^{\lambda_5}, \tag{15}$$

where X_{kr} and Q_{kr} are the current and proposed intake levels, and $\hat{\beta}_{mr}$ represents the nutrient's effect on concentration m from the personal model $\hat{G}_k$. The preference error is calculated as the relative difference between intake levels, scaled by the inverse of the nutrient effect. This scaling prioritizes minimal changes for nutrients with weaker effects while allowing greater adjustments for those with stronger effects, balancing dietary modification and concentration targets.

To ensure consistency, the error is normalized by the number of nutrients (R) and the concentrations (M), making it independent of the size of the data set or the configuration of the model. The hyperparameter λ_5 controls the sensitivity, penalizing larger deviations more heavily to enforce minimal changes to the current diet.

The preference component p_{u} maximizes when the proposed intake Q_r aligns closely with the preference. This is modeled using an exponential distribution:

$$p_{\mathrm{u}}(Q_{kr}|\hat{G}_k, \lambda_k, \lambda_5) = \lambda_k \exp\left(-\lambda_k \, u(\hat{G}_k, \, \lambda_5)\right), \tag{16}$$

where λ_5 globally shapes the preference for minimal changes, while λ_k adapts the scale for each individual k, as described in the next section.

Determining a Safe Limit of Personal Preference Strength

As the proposed intake levels deviate further from the current levels, the likelihood under the exponential distribution decreases, approaching zero. To ensure the model remains well-defined, the likelihood must stay above zero and, in practice, remain above a small threshold ϵ, defined by the computer's precision, to maintain numerical stability.

We can calculate a safe threshold for the slope of the exponential distribution in Eq. (16) that still results in the likelihood ϵ at the maximum possible preference error. First, we calculate the maximum error e by comparing the current diet to the intake limits of each nutrient, using the same logic as in the preference Eq. (16):

$$e = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{1}{R} \sum_{r=1}^{R} \max\left(\left|X_r - Q_r^{\min}\right|, |X_r - Q_r^{\max}|\right) \cdot \frac{1}{1 + \left|\hat{\beta}_{mr}\right|} \right)^{\lambda_5} \tag{17}$$

where $Q_r^{\min}$ and $Q_r^{\max}$ are the minimum and maximum recommended intake levels for nutrient r. Then, using this maximum error e, we solve for the highest safe limit of λ_k, which ensures that the likelihood remains above ϵ:

$$\lambda_k^{\mathrm{safe}} \exp\left(-\lambda_k^{\mathrm{safe}} \cdot e\right) = \epsilon \quad \Longrightarrow \quad \lambda_k^{\mathrm{safe}} = -\frac{W(-e \cdot \epsilon)}{e} \tag{18}$$

where W is the principal branch of the Lambert W function (Corless et al. (1996)), applicable when the argument $-e \cdot \epsilon \geq -1/e$, or equivalently, $e \leq 1/\epsilon$.

Larger values of λ_k may still be preferred to emphasize personal preferences, though they can introduce computational challenges if diet proposals approach intake limits. To provide flexibility, we introduce an adjustable hyperparameter λ_3 can add weight above the minimum level with

$$\lambda_k = \lambda_k^{\min} + \frac{\lambda_3}{e} \tag{19}$$

where the λ_3 can be added directly to the minimum level as the maximum error e is already normalized by the number of modeled concentrations (M) and nutrients (R). This keeps the interpretation of λ_3 consistent across different scenarios.

If the current nutrient intake X_r falls outside the range $[Q_r^{\min}, Q_r^{\max}]$, the preference error will exceed a threshold, as the most preferred intake cannot be achieved. Strict personal preferences also narrow recommendations around the current diet, potentially overlooking better dietary options that could improve nutrient concentrations or align better with health goals if more flexibility were allowed.

Finally, Table 3 summarizes all hyperparameters used in the inference model, including their purposes and guidelines for selecting the appropriate values.

Table 3: Summary of hyperparameters used in the inference model with their default values for the analyzed data sets.

	Purpose	**Value description**	**Defaults**
λ_1	Steepness of concentration limits (Eq. 11)	Moderately high value ensures steep transitions at concentration limits; lower to assign non-zero probabilities to initial proposals.	8
λ_2	Concentration limits requirement (Eq. 12)	High value enforces strict adherence to concentration limits; adjust to relax or tighten requirement.	15
λ_3	Personal preference strength (Eq. 19)	Values over 1 increase emphasis on personal preference λ_k; adjust to balance preference with flexibility.	30 (10-40)
λ_4	Transition steepness between mixture components (Eq. 14)	High value provides sharp transition between components; typically suitable.	30
λ_5	Preference error penalization (Eq. 16)	Values over 1 add penalization of deviations from current diet; adjust with λ_3 for desired balance.	2.0 (1.0-2.0)
λ_6	General recommendations weight (Eq. 6)	Higher values weight recommendations toward general guidelines, steering away from extreme intakes.	10

4 Implementation

Hierarchical models were implemented to estimate nutritional effects and infer optimal nutrient intake using the probabilistic programming language *Stan* (Carpenter et al. (2017)). The Stan implementations for the Sysdimet and Dialysis nutritional effect models are provided in Supplemental Listings S1 and S2 (Turkia et al. (2025)), while the diet inference model in Supplemental Listing S3 (Turkia et al. (2025)) is adaptable to both datasets and beyond.

The Sysdimet model relies on the personal-level linear predictor from Eq. (2), while the Dialysis model includes additional treatment-level grouping to capture variability between treatments. This extension introduces a group effect term and a covariance matrix, allowing the model to account for treatment-specific effects and improve accuracy.

After fitting the hierarchical models, the personal effects of nutrient and current intake estimates were extracted for each patient, following the previously described diet inference process. These were stored as graphical models in GraphML files, enabling efficient sensitivity analyses by adjusting hyperparameters without rerunning posterior inference. Data management and analysis were performed in R, with personal models constructed using the *iGraph* package (Csardi and Nepusz (2006)).

The implementation emphasizes modularity, making the inference model adaptable to various sets of nutrients, dietary restrictions, and personal preferences. The preference error, defined in Eq. (16), was implemented as an adjustable function to maintain numerical stability considering the maximum preference error in Eq. (17).

All figures and posterior analyzes presented in the manuscript and the supplementary material were generated directly from the inference results, ensuring full reproducibility with the provided code.

5 Data Application

We applied the proposed method to infer personalized diet recommendations for patients in the Sysdimet and Dialysis datasets, presenting the results according to the Bayesian Analysis Reporting Guidelines (BARG, Kruschke (2021)).

5.1 Model Evaluation and Posterior Analysis

Posterior inference of the hierarchical effects of nutrients was performed in Stan, drawing 2000 samples from four MCMC chains using the No-U-Turn Sampler (NUTS) algorithm (Homan and Gelman (2014)). Convergence was assessed using rank-based R-hat metrics (Vehtari et al. (2021)), where values below 1.05 indicate well-mixed chains. For the Sysdimet model, the maximum R-hat was 1.014 (mean 1.00), with effective sample sizes (ESS) ranging from 250 to 12,850 (mean 4555). Stan's guidelines (Stan (2024)) recommend an ESS greater than 100 for reliable estimates, while BARG suggests an

ESS greater than 10,000 for high-density regions, indicating reliability for both bulk and tail estimates. No divergent transitions were observed.

For the Dialysis model, the mean R-hat was 1.01 (maximum 1.07), with ESS values ranging from 649 to 8454. Although the maximum R-hat approached the acceptance threshold, visual posterior predictive checks in Supplemental Figs. S1 and S2 (Turkia et al. (2025)) confirmed unbiased fits, and the estimates were deemed suitable.

The normalized root mean square error (NRMSE) evaluations in Supplementary Table S1 (Turkia et al. (2025)) further demonstrated the accuracy of the prediction. In the Sysdimet dataset, HDL-C, LDL-C and total cholesterol had NRMSE values below 0.05, with glucose showing the best fit at 0.03. Insulin predictions were slightly higher at 0.12. In the Dialysis dataset, potassium and albumin had moderate NRMSE values (0.15 and 0.18), while phosphate had a higher value of 0.31.

5.2 Diverse Personal Effects of Nutrients

Hierarchical models revealed substantial individual variability in the effects of nutrients, which we analyzed by building personalized graphical models for each patient. Supplemental Tables S1 and S2 (Turkia et al. (2025)) summarize the 20 strongest personal effects for both datasets, highlighting the nutrients most influential in balancing concentrations.

To illustrate the effectiveness of personalized dietary recommendations, we focused on patients whose current concentrations exceeded the recommended limits. Table 2 highlights these problematic limits in bold, highlighting their role in concentration control. In the Sysdimet study, the main challenge was to reduce glucose and insulin concentrations within the target ranges. For insulin, alpha-linolenic acid (ALA) exhibited the strongest decreasing effect, with minimal variation between individuals (ranging from -7.90 to -8.02), along with a minor increasing effect on glucose (0.37 to 0.38). In contrast, linoleic acid (LA) strongly increased insulin concentrations (7.45 to 7.56) but slightly decreased glucose (-0.39 to -0.40). Vitamin C demonstrated a notable variability in its insulin effects, ranging from a decrease of -0.66 to an increase of 3.17 in normalized magnitude.

In the Dialysis study, most patients faced difficulties in achieving albumin concentrations above the lower limit. Supplemental Table S2 (Turkia et al. (2025)) summarizes the personal effects of various nutrients, revealing substantial variability. For example, salt intake showed effects on albumin ranging from -2.65 to 3.66, while polyunsaturated fat (PUFA) ranged from -1.54 to 2.31. Such opposing effects highlight the complexity of balancing concentrations, underscoring the importance of statistical inference to navigate these challenges.

5.3 Personally Inferred Dietary Recommendations

The diet inference results demonstrate how the theory described in the Methods section translates into actionable personalized recommendations. The prior nutrient intake was

based on the ranges in Table 1, while the concentration limits were defined by the healthy reference ranges in Table 2. The model, implemented as a mixture distribution (Eq. 13), integrates concentration constraints with personal preferences to achieve balanced nutrient intakes.

To validate the inferred recommendations, we ensured that predicted concentrations fell within the targeted limits. Figure 4 illustrates the logarithmic probabilities of the expected concentrations $\hat{\mu}_{kmi}$, showing how the model initially prioritizes the concentration constraints. As the preference component gains weight, the distribution shifts toward diets more aligned with individual preferences. These results demonstrate the successful application of the theoretical framework to real-world datasets.

For five patients from the Sysdimet dataset, Fig. 4 shows how personalized diets brought concentrations within healthy ranges. Dashed vertical lines represent current concentrations, while the solid lines indicate predicted concentrations after dietary adjustments. This outcome was consistent between all patients in both datasets, as detailed in Supplementary Figs. S3 and S5 (Turkia et al. (2025)).

Table 4 summarizes the nutrient intakes corresponding to the concentrations in Fig. 4. The expected intakes reflect the most likely values within the joint recommendation distribution, with 90% credible intervals (CI) indicating viable dietary solutions. Narrow credible intervals suggest nutrients are tightly constrained to meet concentration targets, while wider intervals highlight nutrients with greater flexibility but critical roles in achieving balance.

For example, EPA for patient 104 increased considerably from 0.2 to 2.4 g/d (CI: 2.2 to 2.5 g/d), highlighting its importance for concentration control. Similarly, vitamin C rose from 161 to 840 mg/day (CI: 432 to 996 mg/day), indicating its critical role. In contrast, the protein for patient 1 showed a narrow interval (18.3 to 20.0 g/d) with a minimal change, reflecting its tightly regulated impact.

These results are elaborated in the supplemental material (Turkia et al. (2025)). Supplemental Table S3 provides intake results for all patients in the Sysdimet study, with similar findings for the dialysis cohort. In the dialysis cohort, 35 of 37 patients initially had concentrations outside the normal ranges. Supplementary Fig. S4 illustrates examples of logarithmic probability accumulation for renal patients, while Supplementary Table S6 presents the corresponding intake distributions.

6 Discussion

We have developed a joint probabilistic model that describes the relationship between nutrient intake and a set of concentrations, extending our previous work by enabling the inference of dietary patterns rather than a limited set of nutrients. Implementing this model as a Bayesian mixture framework provides a flexible and interpretable method for addressing this complexity. Simulations based on data from two nutritional studies demonstrate that this approach can generate personalized dietary recommendations

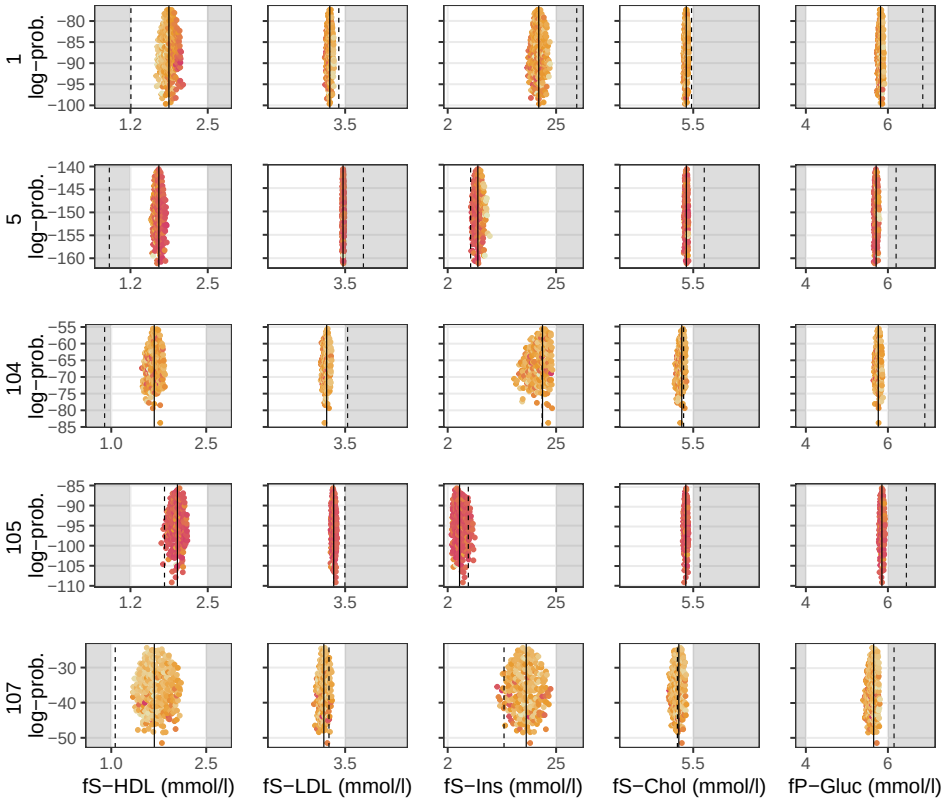

Figure 4: Example of five patients from the control cohort of the Sysdimet study, whose current concentrations deviate from normal ranges but are improved with personalized diet recommendations. The white areas represent the normal concentration ranges, while the dashed vertical lines mark the patients' current concentrations, some of which fall within the gray areas. Each point within the red and yellow distributions represents an inferred diet proposal, with brighter yellow shades indicating closer proximity to the patient's existing diet, which serves as a preference in the recommendation model. The solid lines indicate the concentration levels corresponding to the expected value of the joint multivariate recommendation distribution, where the nutrient intake and concentration predictions are jointly optimized. The diet side of this joint distribution reflects the personalized diet recommendations, ensuring that the expected concentrations are within healthy ranges. The figure was created using the ggplot2 package for R (v 3.3.5, https://ggplot2.tidyverse.org).

Table 4: Personally inferred intake recommendations for the five example patients from Sysdimet study in Fig. 4 whose current concentrations were out of the normal ranges.

	Sucrose	Protein	Carboh.	SFA	MUFA	PUFA	Fiber	Vit. D	Linol. acid	α-linol. acid	Chol.	Folic acid	Vit. C	EPA	DHA
1	8.7→ **1.3** [0.1;3.4]	20.1→ 19.6 [18.3;20.0]	51.3→ 55.4 [48.4;59.9]	11.1→ 9.2 [8.0;10.0]	7.2→ **19.2** [16.5;20.0]	3.6→ **9.6** [8.6;10.0]	26.4→ 21.3 [20.0;24.9]	6.1→ **98.0** [93.0;99.9]	2.7→ **6.1** [3.1;8.8]	0.6→ **2.5** [0.9;3.8]	232→ **10** [0;37]	373→ 290 [125;461]	199→ **78** [75;87]	0.0→ **1.3** [1.0;1.6]	0.1→ **0.0** [0.0;0.1]
5	7.7→ **0.1** [0.0;0.1]	17.1→ 19.9 [19.8;20.0]	53.9→ 59.0 [57.5;59.9]	9.1→ **7.7** [7.1;8.4]	10.0→ **19.9** [19.7;20.0]	4.8→ 5.2 [5.0;5.5]	21.0→ 20.2 [20.0;20.7]	4.8→ **99.1** [97.0;99.9]	3.7→ **2.6** [2.5;2.9]	0.9→ **5.0** [5.0;5.0]	163→ **3** [0;7]	226→ **991** [976;999]	86→ 87 [76;121]	0.0→ **2.0** [1.9;2.0]	0.0→ **0.0** [0.0;0.0]
104	4.4→ **2.6** [0.2;8.5]	17.8→ 17.3 [13.6;19.9]	48.7→ 53.9 [45.8;59.8]	12.0→ **8.8** [7.4;10.0]	10.1→ 11.6 [10.0;15.2]	5.1→ **8.2** [5.9;9.9]	24.0→ 26.6 [22.0;32.8]	8.3→ **95.5** [85.2;99.9]	3.8→ **6.5** [3.3;8.9]	0.5→ **1.0** [0.5;2.4]	346→ **34** [1;131]	282→ 347 [177;584]	161→ **840** [432;996]	0.2→ **2.4** [2.2;2.5]	0.6→ **0.1** [0.0;0.2]
105	4.5→ **0.8** [0.0;2.4]	21.4→ 19.4 [17.8;20.0]	49.6→ 58.7 [55.2;60.0]	10.8→ 9.5 [8.4;10.0]	7.3→ **19.6** [18.7;20.0]	4.9→ **9.3** [7.2;10.0]	19.8→ **46.0** [28.6;54.8]	8.3→ **98.6** [95.2;100.0]	1.7→ **6.3** [3.3;8.8]	0.3→ **2.3** [0.9;3.3]	210→ **11** [0;36]	311→ **968** [889;999]	219→ **83** [75;104]	0.0→ **0.9** [0.7;1.1]	0.1→ **0.0** [0.0;0.1]
107	7.9→ **3.5** [0.3;9.3]	13.7→ 15.0 [10.5;19.6]	40.6→ 50.1 [41.2;59.0]	13.7→ **8.6** [7.2;9.9]	15.5→ 15.1 [10.7;19.5]	6.4→ 7.4 [5.4;9.8]	14.9→ **36.1** [22.4;54.0]	9.9→ **84.5** [57.4;99.3]	3.1→ **5.9** [2.9;8.7]	0.8→ **2.3** [0.6;3.9]	277→ **139** [13;270]	239→ 279 [132;452]	120→ 133 [78;252]	0.1→ **1.5** [0.9;2.2]	0.2→ 0.2 [0.0;0.8]

The table presents the current levels of nutrients intake from patients, the inferred recommendations and the intervals 90% of the recommendations. The increase in current intake is highlighted in red, and the decrease is highlighted in blue. Bold font highlights over 30% change in the current intake.

capable of restoring unhealthy nutrient concentrations with minimal individual dietary modifications.

While our computational results are promising, validating these personalized dietary recommendations would require clinical interventions, which have not yet been conducted. The absence of ground truth data makes it challenging to compare directly with other recommendation methods. To establish the validity of our approach, we evaluated the alignment of our model's nutrient effect estimates and intake recommendations with established findings in the nutrition literature, finding a good agreement with current nutritional understanding.

6.1 Evaluation of the Inferred Recommendations

Supplementary Table S8 (Turkia et al. (2025)) compares our inferred nutrient effects and intake recommendations from both data sets with existing nutrition literature. Our analysis identifies two distinct nutrient groups: those with consistent directional effects (increasing or decreasing concentrations) that vary only in magnitude across individuals and those with mixed effects that vary in direction among individuals. Although the literature supports our findings for nutrients with consistent effects, the mixed effects are less documented in existing research. The model's alignment with well-established nutritional effects supports its validity, and identifying personally varying effects may reflect limitations in traditional statistical approaches rather than model shortcomings.

Our results demonstrate that the Bayesian framework offers advantages in personalized nutrition modeling. The transparency and interpretability of our customized Bayesian model – key attributes highlighted by Gleason and Harris (Gleason and Harris (2019)) – are evident in our analysis of the contributions of nutrients to plasma concentrations. Figures 5 and 6 illustrate these contributions,

ranked by impact magnitude. The figures show how recommended nutrient intakes adjust concentration levels from their baseline values (set by fixed predictors) to achieve expected values according to Eq. (2), providing clear insight into the model's decision-making process.

The resulting concentration levels reflect complex interactions among the effects of nutrients, personal baselines, and current dietary patterns, rather than simple nutrient-concentration relationships. In the Sysdimet cohort (Fig. 5), the model identified key contributors to insulin concentration, including EPA, folate, alpha-linolenic acid, linoleic acid, and vitamin C. EPA exhibited substantial individual variability with strong increasing effects in some patients, while alpha-linolenic acid showed a more consistent and moderate positive impact. Linoleic acid consistently decreased insulin concentration, whereas vitamin C displayed varying effects, ranging from decreases to increases across individuals. Detailed numerical evidence for these contributions is provided in Supplemental Tables S5 and S6 (Turkia et al. (2025)). In particular, these computationally inferred effects, derived from observational Sysdimet data, align closely with the findings confirmed in the intervention cohorts of the same study (Lankinen et al. (2011)). This alignment highlights the ability of our model to extract clinically relevant insights from observational data alone.

Protein is vital for renal patients due to its positive effect on albumin (0.69 to 2.00) and consistent insulin-lowering impact (-0.50 to -0.84). Similarly, sucrose, which increases albumin (0.70 to 2.65) but decreases potassium and insulin, is typically reduced to balance these concentrations. Carbohydrates have moderate effects, supporting adjustments towards a 50% E intake to reduce insulin while increasing albumin and potassium levels (Kiens and Richter (1996); Leon et al. (2001)). Nutrients with less pronounced effects, such as fiber and water, remain near general recommendations unless their current intake exceeds the suggested ranges.

Our model provides nuanced recommendations for nutrients with complex effects, illustrated by MUFA and PUFA. MUFA exhibits varying individual-specific impacts on insulin (-0.83 to 1.31) and potassium (-0.20 to 0.34). However, the model recommends increasing intake towards 15 E%, which aligns with current dietary guidelines (Blomhoff et al. (2023)). Similarly, while PUFA generally lowers fasting plasma insulin concentration, its effects on albumin vary substantially between individuals (-1.54 to 2.31), highlighting the importance of personalized recommendations.

The model successfully balances competing nutrient effects to achieve desired concentration levels. This is evident in the case of patient 104 (marked with a dashed line in Fig. 5), where vitamin C and EPA reduce fasting plasma glucose concentration. At the same time, linoleic acid counterbalances the insulin-increasing effect of EPA.

The smooth and differentiable concentration limits in our mixture model (Eq. 13) allow meaningful recommendations even when optimal solutions are not achievable. For example, in Fig. 5, although unable to bring patient 121's high insulin concentrations within optimal limits through dietary modifications alone as detailed in Supplementary Table S3 (Turkia et al. (2025)), the model still provides valid dietary recommendations by proceeding with the inference.

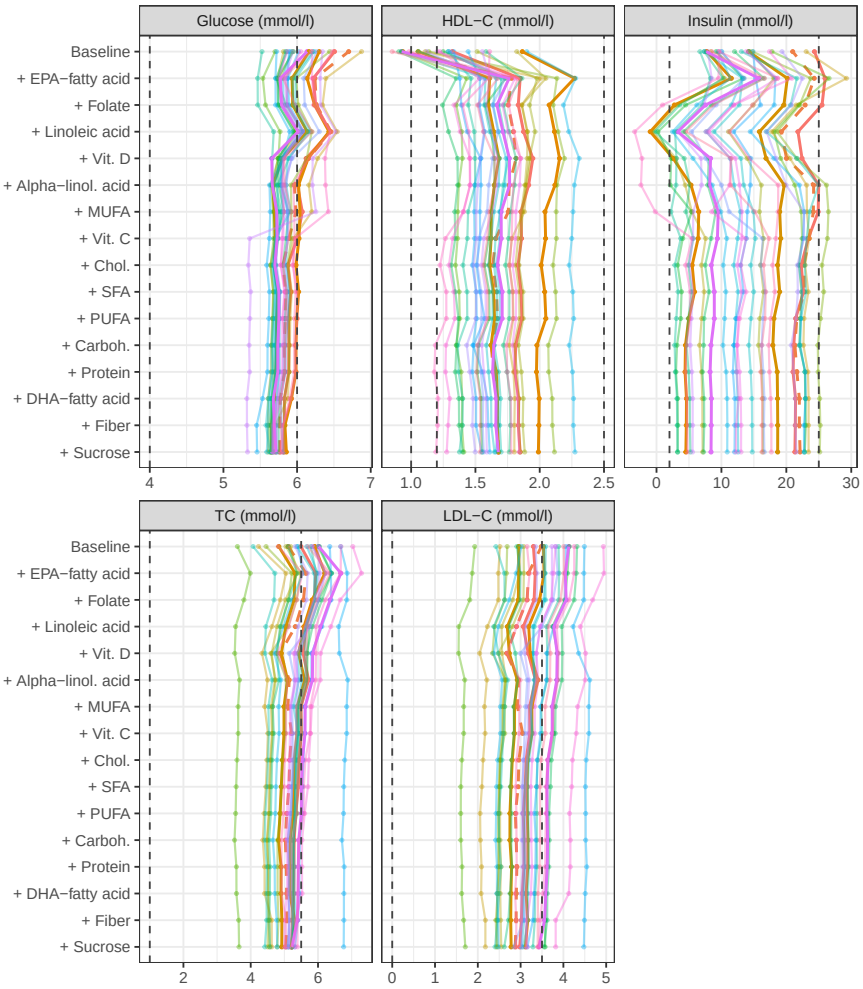

Figure 5: Figure shows personal compositions of concentrations for Sysdimet data set patients, starting with baseline concentrations and continuing with the contributions of each inferred nutrient intake. At the bottom, the final concentrations are predicted to reside within the targeted limits, as the dashed lines show. Nutrients are arranged in decreasing order of average contribution. Each colored line represents an accumulation of concentration for one patient. The thicker lines correspond to the example patients in Fig. 4 and Table 4. The figure is plotted with ggplot2 package for R language (v 3.3.5, https://ggplot2.tidyverse.org).

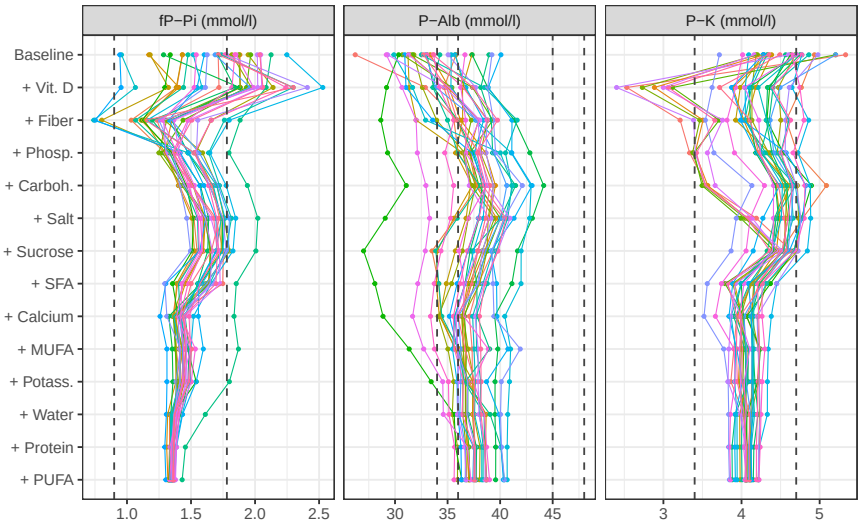

Figure 6: Figure shows personal compositions of concentrations for Dialysis data set patients starting with baseline concentrations and continuing with the contributions of each inferred nutrient intake. At the bottom, the final concentrations are predicted to reside within the targeted limits, as the dashed lines show. Nutrients are arranged in decreasing order of average contribution. The figure is plotted with ggplot2 package for R language (v 3.3.5, https://ggplot2.tidyverse.org).

6.2 Limitations of the Method

Although our Bayesian model provides a robust framework for balancing nutrient intake with plasma concentration targets, several limitations should be acknowledged. First, the model assumes that changes in plasma concentrations are directly attributable to nutrient intake, which may oversimplify complex physiological processes and overlook non-dietary factors. The observed variations in individual dietary responses, captured as latent variables, likely reflect differences in genetics, metabolism, lifestyle, and environmental factors (Boorsma et al. (2017)).

To simplify the analysis of limited nutritional data, the model assumes that the expected concentrations are a linear combination of the intake of nutrients and the effects of nutrients. Although actual biological reactions are often non-linear and more complex (Martini et al. (2017)), this linear additive assumption applies only to the expected concentration values, allowing for a straightforward interpretation of nutritional effects. The model accommodates natural variability in observed concentrations through a Gamma distribution, maintaining analytical simplicity while reflecting inherent biological fluctuations.

Flexibility in the model is achieved through several hyperparameters that control concentration limits and the weighting of personal preferences (Table 3). Although these hyperparameters improve adaptability, they also require careful tuning to ensure an appropriate balance between meeting concentration targets and generating realistic dietary recommendations. Sensitivity analysis is essential to select suitable values.

In cases where concentration limits cannot be fully met, the model can produce recommendations that are computationally effective but not biologically optimal. For example, controversially high intake suggestions for vitamins C and D were observed in subjects S104, S108, S12, S48, S64, and S69 (see Fig. S4). In these cases, the estimated insulin-lowering effect of vitamin C was used as a last resort for individuals struggling to meet the upper glucose and insulin limits. Although these recommendations are computationally justified, alternative clinical strategies may be more appropriate.

6.3 Reliability of the Analysis

Detailed diagnostics of the model support the reliability of the analysis. The Sysdimet and Dialysis models demonstrated effective convergence, with R-hat values below 1.05 and high effective sample sizes across parameters. However, the relatively small number of observations in both datasets contributes to wider credible intervals for some nutrient effect coefficients, reflecting greater uncertainty in these estimates. This is particularly notable for nutrients with high individual variability, such as EPA and vitamin C, where credible intervals remain broad despite robust convergence metrics.

Although Bayesian inference offers significant advantages, it is computationally demanding. The iterative nature of sampling in Bayesian estimation increases computational costs, particularly as the number of covariates and model complexity grows. Variational approximations could help mitigate these challenges by improving efficiency, though at the cost of approximation. Despite these trade-offs, the Bayesian

framework remains a powerful tool for personalized nutrition, especially when prioritizing long-term health metrics and individual preferences.

7 Conclusions

Our results demonstrate that the proposed Bayesian inference model effectively manages the complex interplay of personal concentration baselines, competing nutritional effects, and current dietary intake, enabling personalized dietary strategies that address individual health needs. By balancing concentration constraints and personal preferences, the model provides actionable recommendations tailored to individual responses, using plasma concentrations as indicators of overall health status.

We successfully replicated, using observational data, the health benefits of omega-3 fatty acids from fish and vitamin C from flavonoids in improving glucose metabolism and lipid profiles – findings previously confirmed in the Sysdimet study (Lankinen et al. (2011)) through nutritional interventions. This alignment highlights the potential of observation-driven dietary guidance as a practical and efficient alternative to resource-intensive intervention studies, bridging the gap between research and real-world applications. Furthermore, the Bayesian framework's ability to quantify uncertainty ensures robust, evidence-based recommendations, advancing personalized nutrition as a promising approach to improving public health.

Supplementary Material

Supplemental document. Document contains the supplemental code listings, figures, and tables referenced in the main article, organized to correspond with its sections.

References

Ben-Yacov, O., Godneva, A., Rein, M., Shilo, S., Kolobkov, D., Koren, N., Cohen Dolev, N., Travinsky Shmul, T., Wolf, B. C., Kosower, N., Sagiv, K., Lotan-Pompan, M., Zmora, N., Weinberger, A., Elinav, E., and Segal, E. (2021). "Personalized postprandial glucose response-targeting diet versus Mediterranean diet for glycemic control in prediabetes." *Diabetes Care*, 44(9): 1980–1991. 2

Blomhoff, R., Andersen, R., Arnesen, E., Christensen, J., Eneroth, H., Erkkola, M., Gudanaviciene, I., Halldorsson, T., Høyer-Lund, A., Lemming, E., Meltzer, H., Pitsi, T., Siksna, I., Thorsdottir, I., and Trolle, E. (2023). *Nordic Nutrition Recommendations 2023*. Nordisk Ministerråd. 3, 5, 21

Boorsma, A., van Someren, E., de Hoogh, I., Hogenelst, K., van Erk, M., Wopereis, S., Rouhani-Rankouhi, T., van den Broek, T., Pasman, W., van Ommen, B., and Anthony, J. C. (2017). "Systems biology of personalized nutrition." *Nutrition Reviews*, 75(8): 579–599.
URL https://doi.org/10.1093/nutrit/nux029 2, 24

Bush, C. L., Blumberg, J. B., El-Sohemy, A., Minich, D. M., Ordovás, J. M., Reed, D. G., and Behm, V. A. Y. (2020). "Toward the definition of personalized nutrition: A proposal by the American nutrition association." *J. Am. Coll. Nutr.*, 39(1): 5–15. 2

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software, Articles*, 76(1): 1–32.
URL https://www.jstatsoft.org/v076/i01 16

Cecil, J. E. and Barton, K. L. (2020). "Inter-individual differences in the nutrition response: from research to recommendations." *Proceedings of the Nutrition Society*, 79(2): 171–173. 2

Celis-Morales, C., Livingstone, K. M., Marsaux, C. F. M., Macready, A. L., Fallaize, R., O'Donovan, C. B., Woolhead, C., Forster, H., Walsh, M. C., Navas-Carretero, S., San-Cristobal, R., Tsirigoti, L., Lambrinou, C. P., Mavrogianni, C., Moschonis, G., Kolossa, S., Hallmann, J., Godlewska, M., Surwiłło, A., Traczyk, I., Drevon, C. A., Bouwman, J., van Ommen, B., Grimaldi, K., Parnell, L. D., Matthews, J. N. S., Manios, Y., Daniel, H., Martinez, J. A., Lovegrove, J. A., Gibney, E. R., Brennan, L., Saris, W. H. M., Gibney, M., and Mathers;, J. C. (2016). "Effect of personalized nutrition on health-related behaviour change: evidence from the Food4me European randomized controlled trial." *Int. J. Epidemiol.*, dyw186. 2

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996).

"On the LambertW function." *Advances in Computational Mathematics*, 5(1): 329–359.
URL https://doi.org/10.1007/BF02124750 15

Csardi, G. and Nepusz, T. (2006). "The igraph software package for complex network research." *InterJournal*, Complex Systems: 1695.
URL https://igraph.org 16

Duodecim, F. M. S. (2024). "Current Care Guidelines." https://www.kaypahoito.fi. Accessed: 1.2.2025. 6, 7

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
URL https://books.google.fi/books?id=ZXL6AQAAQBAJ 11

Gleason, P. M. and Harris, J. E. (2019). "The Bayesian Approach to Decision Making and Analysis in Nutrition Research and Practice." *Journal of the Academy of Nutrition and Dietetics*, 119(12): 1993–2003.
URL https://www.sciencedirect.com/science/article/pii/S2212267219308640 2, 20

Homan, M. D. and Gelman, A. (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.*, 15(1): 1593–1623. 16

Ikizler, T. A., Burrowes, J. D., Byham-Gray, L. D., Campbell, K. L., Carrero, J.-J., Chan, W., Fouque, D., Friedman, A. N., Ghaddar, S., Goldstein-Fuchs, D. J., Kaysen, G. A., Kopple, J. D., Teta, D., Yee-Moon Wang, A., and Cuppari, L. (2020). "KDOQI Clinical Practice Guideline for nutrition in CKD: 2020 update." *Am. J. Kidney Dis.*, 76(3 Suppl 1): S1–S107. 2

Kiens, B. and Richter, E. (1996). "Types of carbohydrate in an ordinary diet affect insulin action and muscle substrates in humans." *The American journal of clinical nutrition*, 63 1: 47–53. 21

Kirk, D., Catal, C., and Tekinerdogan, B. (2021). "Precision nutrition: A systematic literature review." *Comput. Biol. Med.*, 133(104365): 104365. 2

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press. 7

Kruschke, J. K. (2021). "Bayesian Analysis Reporting Guidelines." *Nature Human Behaviour*, 5(10): 1282–1291.
URL https://doi.org/10.1038/s41562-021-01177-7 3, 16

Lankinen, M., Schwab, U., Kolehmainen, M., Paananen, J., Poutanen, K., Mykkanen, H., Seppanen-Laakso, T., Gylling, H., Uusitupa, M., and Oresic, M. (2011). "Whole grain products, fish and bilberries alter glucose and lipid metabolism in a randomized, controlled trial: the Sysdimet study." *PLoS ONE*, 6(8): e22646. 5, 21, 25

Leon, J., Majerle, A., Soinski, J., Kushner, I., Ohri-Vachaspati, P., and Sehgal, A.

(2001). "Can a nutrition intervention improve albumin levels among hemodialysis patients? A pilot study." *Journal of Renal Nutrition*, 11(1): 9–15. 21

Limpert, E. and Stahel, W. A. (2011). "Problems with Using the Normal Distribution – and Ways to Improve Quality and Efficiency of Data Analysis." *PLOS ONE*, 6(7): 1–8.
URL https://doi.org/10.1371/journal.pone.0021403 9

Livingstone, K. M., Celis-Morales, C., Navas-Carretero, S., San-Cristobal, R., Forster, H., Woolhead, C., O'Donovan, C. B., Moschonis, G., Manios, Y., Traczyk, I., Gundersen, T. E., Drevon, C. A., Marsaux, C. F. M., Fallaize, R., Macready, A. L., Daniel, H., Saris, W. H. M., Lovegrove, J. A., Gibney, M., Gibney, E. R., Walsh, M., Brennan, L., Martinez, J. A., and Mathers, J. C. (2020). "Characteristics of participants who benefit most from personalised nutrition: findings from the pan-European Food4Me randomised controlled trial." *Br. J. Nutr.*, 123(12): 1396–1405. 2

Martini, F. H., Nath, J. L., and Bartholomew, E. F. (2017). *Fundamentals of Anatomy and Physiology*. Boston: Pearson, 11th edition. Chapter 21: The Cardiovascular System: Blood Vessels and Circulation. 2, 24

Mehtatalo, L. and Lappi, J. (2020). *Biometry for Forestry and Environmental Data: With Examples in R (1st ed.)*.. Chapman and Hall/CRC.
URL https://doi.org/10.1201/9780429173462 9

Rein, M., Ben-Yacov, O., Godneva, A., Shilo, S., Zmora, N., Kolobkov, D., Cohen-Dolev, N., Wolf, B.-C., Kosower, N., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Zelber-Sagi, S., Elinav, E., and Segal, E. (2022). "Effects of personalized diets by prediction of glycemic responses on glycemic control and metabolic health in newly diagnosed T2DM: a randomized dietary intervention pilot trial." *BMC Med.*, 20(1): 56. 2

Stan, S. D. T. (2024). *Stan Modeling Language Users Guide and Reference Manual*.
URL https://mc-stan.org 16

Turkia, J., Mehtätalo, L., Schwab, U., and Hautamäki, V. (2021). "Mixed-effect Bayesian network reveals personal effects of nutrition." *Scientific Reports*, 11(1): 12016.
URL https://doi.org/10.1038/s41598-021-91437-3 2, 4

Turkia, J., Schwab, U., and Hautamäki, V. (2024). "Inferring personal intake recommendations of phosphorous and potassium for end-stage renal failure patients by simulating with Bayesian hierarchical multivariate model." *PLOS ONE*, 19(2): 1–25.
URL https://doi.org/10.1371/journal.pone.0291153 3, 4, 5, 9

— (2025). "Supplement to "Personal Diet Inference through Mixture of Concentration Constraints and Intake Preferences."." 16, 17, 18, 20, 21

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). "Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence

of MCMC (with Discussion).” *Bayesian Analysis*, 16(2): 667 – 718. URL https://doi.org/10.1214/20-BA1221 16

World Health Organization (2024). “Nutrition.” Accessed: 2024-11-10. URL https://www.who.int/health-topics/nutrition 2

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J. A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., Bikovsky, R., Halpern, Z., Elinav, E., and Segal, E. (2015). “Personalized nutrition by prediction of glycemic responses.” *Cell*, 163(5): 1079–1094. 2