data

MDPI

Article

Statistical Approach in Personalized Nutrition Exemplified by Reanalysis of Public Datasets

Paola G. Ferrario [1,*], Maik Döring [2] and Christian Ritz [3]

[1] Department of Physiology and Biochemistry of Nutrition, Max Rubner-Institut, 76131 Karlsruhe, Germany
[2] National Reference Centre for Authentic Food, Max Rubner-Institut, 95326 Kulmbach, Germany; maik.doering@mri.bund.de
[3] National Institute of Public Health, University of Southern Denmark, 1455 Copenhagen K, Denmark; ritz@sdu.dk
* Correspondence: paola.ferrario@mri.bund.de

Abstract: In clinical nutrition, it is regularly observed that individuals respond differently to a dietary treatment. Personalized nutrition aims to consider such variability in response by delivering personalized nutritional recommendations. Ideally, the optimal treatment for each individual will be selected and then dispensed according to the specific individual's characteristics. The aim of this paper is to discuss and apply existing statistical methods, which can be adequately used in the context of personalized nutrition. We discuss the estimation of individualized treatment rules (ITRs) as we wish to favor one out of two interventions. The applicability of the methods is demonstrated by reusing two public datasets: one in the context of a parallel group design and one in the context of a crossover design. The bias of the estimator of the ITRs underlying parameters is evaluated in a simulation study.

Keywords: personalized nutrition; individualized treatment rules; simulations; real data example; publicly available datasets; parallel group and crossover design

Academic Editor: Jamal Jokar Arsanjani

Received: 17 October 2024
Revised: 23 January 2025
Accepted: 27 January 2025
Published: 30 January 2025

Citation: Ferrario, P.G.; Döring, M.; Ritz, C. Statistical Approach in Personalized Nutrition Exemplified by Reanalysis of Public Datasets. *Data* **2025**, *10*, 18. https://doi.org/10.3390/data10020018

1. Introduction

In nutrition research, heterogeneity in response to dietary interventions has been repeatedly observed. To give just a few examples, heterogeneity in stimulation after caffeine intake [1] has already been established, in addition to heterogeneity in postprandial glycemia [2] in weight loss [3] and in absorption, metabolism, and excretion of nutrients [4]. Personalized nutrition aims to take this heterogeneity into account by providing individuals with different nutritional recommendations, depending on whether or not they are carriers of specific genetic variants, for example. Since the cost as well as the benefits of a dietary intervention might be difficult to predict in advance, it is useful to know how the magnitude of the benefits vary with individual attributes (i.e., by a covariate) (Athey and Imbens [5]).

Standard statistical methods considering and comparing population averages of some parameters of interest may not be adequate for exploiting such heterogeneity in human response to treatments since heterogeneity is often seen as a nuisance parameter.

More appropriate statistical tools in such a personalized context already exist, and they are receiving greater attention: so-called *individualized treatment rules (ITRs)*. ITRs aim to assign treatments, favoring one treatment over alternative treatments, in order to achieve optimal outcomes and can be expressed via the so-called *conditional average treatment effects function D*.

Nevertheless, some work exists in the literature on the estimation of the conditional treatment effects function in question. Most notably, Matsouaka et al. [6] proposed a two-step robust method to derive ITRs and evaluate their values; Kuenzel et al. [7] proposes a unifying framework for many estimators of the function D, using different machine learning techniques; Athey and Imbens [5] propose a partitioning approach in order to find subpopulations that differ in the magnitude of treatment effects, focusing on estimating and making inference for the function D. Moreover, extensions that balance bias and variance inflation, when using regularization, have also been proposed recently by Cai et al. [8], in addition to a constrained Lasso approach of hierarchical interactions for predicting treatment response and analyzing heterogeneity of treatment effects proposed by Du et al. [9].

The main focus of the current paper is to show and discuss the estimation of the conditional average treatment effects function D in a clinical context. Specifically, the contribution of this paper consists in identifying and transferring existing statistical methods, which can be useful in personalized nutrition, in addition to illustrating application to real data after practical implementation of the estimators.

The rest of the paper is organized as follows. Section 2 introduces some notation and assumptions and discusses an estimation of D; Section 3 provides two illustrative examples and includes a simulation study. Section 4 outlines and discusses possible extensions of the framework.

2. Methods

Notation and Assumptions

Let $(\mathbf{X}, Y, G)$, $(\mathbf{X}_1, Y_1, G_1), \ldots, (\mathbf{X}_n, Y_n, G_n)$ be a $(\mathbb{R}^d \times \mathbb{R} \times \{0,1\})$ i.i.d triple, where $\mathbf{X}$ is a baseline covariate vector, Y is the outcome variable, and G a dichotomous variable indicating treatment choice over two: G is equal to 0 for a *standard or control treatment*, and G is equal to 1 for an *experimental treatment*, which are assigned randomly. The dimension of the covariate $\mathbf{X}$ may be relatively large, and the covariate may or may not be continuous.

Throughout the article, we use the notation $Y^{(0)}$, which denotes the potential outcome of Y in case where G was equal to 0 (standard treatment), and we use the notation $Y^{(1)}$, which denotes the potential outcome of Y in case where G was equal to 1 (experimental treatment).

Y, $Y^{(0)}$, and $Y^{(1)}$ are related via the *consistency assumption*, requiring that $Y = GY^{(1)} + (1-G)Y^{(0)}$ [6]. Moreover, we assume that each individual's potential response to treatment does not depend on the treatment assignment mechanism, the treatment received by other individuals, or their potential responses to treatment (*standard stable unit treatment value assumption*) [10]. The assumption of *ignorable treatment assignment* rules out the existence of unobservable factors that affect the treatment selection, meaning that $(Y^{(0)}, Y^{(1)})$ and G are conditional independent with respect to $\mathbf{X}$ (Rosenbaum et al. [11]). Finally, we assume that a larger value of Y is more beneficial, without loss of generality [6].

Let $\mathcal{I}(\mathbf{X})$ be an individualized treatment rule (ITR), i.e., a function $\mathbb{R}^d \to \{0,1\}$ that assigns one of the two treatments to individuals (i.e., 0 for treatment 0 and 1 for treatment 1), given the vector $\mathbf{X}$ of the (observable) covariate.

The final aim is to identify an optimal $\mathcal{I}(\cdot)$, i.e., an optimal treatment assignment, which maximizes individuals' outcomes, depending on some of their characteristics. When the treatment choice is optimized for all individuals, the resulting population average outcome is also optimized. Thus, an optimal ITR is also expected to maximize a population average value function. One way to quantify such individuals' outcomes is by the *population average outcome*, defined as

$$\mathcal{V}_{\mathcal{I}} := \mathbb{E}\Big\{\mathcal{I}(\mathbf{X})Y^{(1)} + (1-\mathcal{I}(\mathbf{X}))Y^{(0)}\Big\}. \tag{1}$$

The population average outcome is optimized as long as treatment assignment is optimized for all individuals. The optimal $\mathcal{I}(\cdot)$ maximizing $\mathcal{V}_{\mathcal{I}}$ is Bayes Rule

$$\mathcal{I}_{Bayes} = argmax_{\mathcal{I}}\mathcal{V}_{\mathcal{I}} \quad \text{with} \quad \mathcal{I}_{Bayes}(\mathbf{X}) = 1\Big\{\mathbb{E}\Big\{\Big(Y^{(1)} - Y^{(0)}\Big)|\mathbf{X}\Big\} \geq 0\Big\} \tag{2}$$

where $1\{\cdot\}$ is the indicator function (compare [6]). Bayes Rule (2) is determinated by a function $\mathbb{R}^d \to \mathbb{R}$, called *conditional average treatment effects function* (CATE), defined as

$$D(\mathbf{x}) := \mathbb{E}\Big\{\Big(Y^{(1)} - Y^{(0)}\Big)|\mathbf{X} = \mathbf{x}\Big\}, \tag{3}$$

which can be expressed as $\mu^{(1)}(\mathbf{x}) - \mu^{(0)}(\mathbf{x}) = \mathbb{E}\Big\{Y^{(1)}|\mathbf{X} = \mathbf{x}\Big\} - \mathbb{E}\Big\{Y^{(0)}|\mathbf{X} = \mathbf{x}\Big\}$. Therefore, $\mu^{(0)}$ and $\mu^{(1)}$ refer to the conditional expected potential outcome with the standard treatment and the experimental treatment, respectively, factorized on the covariate value. Since the distribution of $(\mathbf{X}, Y, G)$ and, therefore, $\mathcal{I}$ and D are typically unknown, we have to estimate them by a sample $(\mathbf{X}_1, Y_1, G_1) \ldots (\mathbf{X}_n, Y_n, G_n)$ of i.i.d. copies of $(\mathbf{X}, Y, G)$. Then, the function D as in Equation (3) can be estimated by

$$\widehat{D}(\mathbf{x}) = \widehat{\mu}^{(1)}(\mathbf{x}) - \widehat{\mu}^{(0)}(\mathbf{x}), \tag{4}$$

where $\widehat{\mu}^{(0)}$ and $\widehat{\mu}^{(1)}$ are estimators of $\mu^{(0)}$ and of $\mu^{(1)}$. If the two true functions $\mu^{(0)}$ and $\mu^{(1)}$ are parallel, their difference would represent a treatment effect independent of $\mathbf{x}$. Otherwise, if $\mu^{(0)}$ and $\mu^{(1)}$ are not parallel, the difference varies with $\mathbf{x}$, which would suggest a distinct benefit by one treatment over the other according to the value of $\mathbf{x}$; this would represent an individualized treatment effect.

The CATE can be given parametrically by means of a linear model, expressed via parameters $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$, i.e., $\widehat{\mu}^{(0)}(\mathbf{x}) = \hat{\boldsymbol{\beta}}_{\mathbf{0}}{}^{(0)} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}{}^{(0)}\mathbf{x}$ and $\widehat{\mu}^{(1)}(\mathbf{x}) = \hat{\boldsymbol{\beta}}_{\mathbf{0}}{}^{(1)} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}{}^{(1)}\mathbf{x}$. Assuming the parameter estimates are asymptotically normal, then, asymptotic confidence intervals of the treatment effect $D(\mathbf{x})$ can be calculated by the delta method. An ITR could be given by assigning the experimental treatment if the lower bound of the confidence interval of the CATE at the observed value of the covariate is bigger than zero. Comparing to Bayes Rule, this is a more conservative approach, which is motivated by a higher potential risk, by applying non-standard treatments. However, an increase in sample size will decrease the size of the confidence interval, and by the consistency of the estimators, the ITR converge for increasing sample size to Bayes Rule.

In the case of a linear regression model with a one-dimensional covariate X the estimator of D becomes

$$\widehat{D}(x) = (\hat{\beta_0}^{(1)} - \hat{\beta_0}^{(0)}) + (\hat{\beta_1}^{(1)} - \hat{\beta_1}^{(0)})x, \tag{5}$$

as in Rosenbaum et al. [11]. In the case of a linear regression model with a two-dimensional covariate vector $\mathbf{X}$, i.e., $\mathbf{X} = (X_1, X_2)^T$ the estimator of D becomes

$$\hat{D}(\mathbf{x}) = (\hat{\beta_0}^{(1)} - \hat{\beta_0}^{(0)}) + (\hat{\beta}_1^{(1)} - \hat{\beta}_1^{(0)})x_1 + (\hat{\beta}_2^{(1)} - \hat{\beta}_2^{(1)})x_2$$

3. Data Examples

3.1. Application to a Trial Concerning Serum Vitamin D Concentration

In order to illustrate the described method, we reconsidered a randomized controlled trial on the effect of two different doses of cholecalciferol on the serum vitamin D concentration

as the primary outcome, see Steenhoff et al. [12], with freely available data at the PLOS ONE repository https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117123.

The study was carried out in 2012 in Gaborone, Botswana. The goal of this study was to investigate whether the two supplementation doses of cholecalciferol of 4000 IU/day versus 7000 IU/day would safely result in an increased serum 25(OH)D concentration. Serum 25(OH)D concentrations were measured at baseline and after 6 and 12 weeks of treatment. To summarize, this trial considered two possible treatments in conjunction with repeated measurements at three time points. The original authors failed to find evidence of heterogeneous treatment effects when using simple linear models. However, the high variance in the serum vitamin D (25(OH)D) concentration, especially after the treatment with the higher dose, suggested that participants reacted heterogeneously to the supplementation.

The data (in the long format) were imported (see Appendix A and Listing 1); the variables of interest were as follows: the serum 25(OH)D concentration at the end of the study (*Serum.25.OH.D.ng.mL*), the two different supplementation doses: 4000 IU/day as standard treatment and 7000 IU/day as experimental treatment (*Vitamin.D3.dose.IU*), the two time points: after 6 and after 12 weeks of treatment (*Time*), and the baseline serum 25(OH)D concentration (*baseline*). Moreover, a new categorical variable was defined (*timXvit*), encoding the combinations of the two supplementation doses and the two time points after 6 and after 12 weeks of treatment. Finally, the adjusting covariates were as follows: sex (*Sex*), age (*Age.yrs*) and weight (*Weight.kg*).

Listing 1. Import and convert the data of the Vitamin D dataset.

```
data.vitd.3nobase<-read.csv2(file.path(dataDir,
                                       "journal.pone.0117123.s006_pgf.csv"))
data.vitd.3nobase[["Vitamin.D3.dose.IU"]] <-
  as.factor(data.vitd.3nobase[["Vitamin.D3.dose.IU"]])
data.vitd.3nobase[["Time"]] <- as.factor(data.vitd.3nobase[["Time.wks"]])
data.vitd.3nobase[["timXvit"]] <- with(data.vitd.3nobase,
                                       interaction(Time, Vitamin.D3.dose.IU))
```

For parametric estimation of the serum 25(OH)D concentration with the baseline serum 25(OH)D concentration as a one-dimensional covariate, we began fitting a treatment-covariate interaction model (Royston and Sauerbrei [13]). For this, the serum 25(OH)D concentration was described by linear mixed models including a specific interaction term: This term combined the categorical variable of the interaction between the doses and the time points (*timXvit*, taking the values 6.4000, 12.4000, 6.7000, and 12.7000, corresponding to the four combinations between time and dosage), together with the continuous variable of the baseline serum vitamin D concentration, *timXvit:baseline*, compare Listing 2. This enables the preservation of as much information as possible, with the baseline serum 25(OH)D concentration unchanged as a continuous variable and avoiding any categorizations for it. The covariate-treatment interaction term was essential for the identification of the individualized treatment effects. The treatment-covariate interaction model implied different intercepts and covariate-specific slope(s) for each time point and each treatment group. Moreover, the model also included individual-specific random effects for capturing between-participant variation ($(1 \mid ID)$) and three adjusting covariates.

Listing 2. The model for the Vitamin D dataset.

```
library(lme4)
pn.lmm <- lmer(Serum.25.OH.D.ng.mL ~ timXvit-1 +
                 timXvit:baseline -baseline + Sex+Age.yrs+Weight.kg+ (1|ID),
               data = data.vitd.3nobase)
```

The output of the calculation consisted of eleven estimated coefficients for the fixed effects, with 95% confidence intervals given in Table 1. With the exception of the interaction term for the time after 12 weeks combined with the higher dose (*timXvit12.7000*), all other intercept and slope parameters were significantly different from zero, with 95% confidence intervals not crossing the zero.

Table 1. Intercept and slope parameters of the fixed effects for the serum 25(OH)D concentration, with baseline serum 25(OH)D concentration as one-dimensional covariate, with adjusting covariates, and with 95% confidence intervals.

	Estimate	**Std. Error**	**2.5%**	**97.5%**
timXvit6.4000	40.90	11.98	18.47	63.32
timXvit12.4000	45.98	11.87	23.75	68.21
timXvit6.7000	26.96	10.97	6.41	47.51
timXvit12.7000	24.39	11.03	3.73	45.05
SexM	−5.02	3.76	−12.07	2.01
Age.yrs	0.37	0.21	−0.03	0.77
Weight.kg	−0.44	0.15	−0.71	−0.16
timXvit6.4000:baseline	0.70	0.29	0.16	1.24
timXvit12.4000:baseline	0.61	0.29	0.07	1.15
timXvit6.7000:baseline	1.25	0.29	0.71	1.79
timXvit12.7000:baseline	1.36	0.29	0.81	1.90

Moreover, the confidence intervals for the estimate $\widehat{D}(x)$ were calculated by the delta method after having named the eleven estimated parameters from a1 to a11 (compare Listing 3 and the first column in Table 1).

Listing 3. Confidence intervals by delta method for $\widehat{D}(x)$ (first dataset).

```
coefVec <- coef(summary(pn.lmm))[, 1]
parmNames <- paste("a", 1:11, sep = "")
names(coefVec) <- parmNames
vcmat.lmm1 <- vcov(pn.lmm)
library(car)
diffFct0 <- function(baseVal)
{
  deltaMethod(coefVec, paste("a4-a2 + (a11-a9)*", baseVal), vcmat.lmm1,
              parameterNames = parmNames, level = 0.95)[, "Estimate"]
}

diffSEFct0 <- function(baseVal)
{
  deltaMethod(coefVec, paste("a4-a2 + (a11-a9)*", baseVal), vcmat.lmm1,
              parameterNames = parmNames, level = 0.95)[, "SE"]
}
```

With this, the parametric estimate $\widehat{D}(x)$ was given by $(24.39 - 45.98) + (1.36 - 0.61) \cdot$ baseline, for any baseline value. The baseline serum vitamin D concentration had a minimum value of 15.30 ng/mL, a first quartile of 28.55 ng/mL, a median of 33.60 ng/mL, a mean of 35.54 ng/mL a third quartile of 41.85 ng/mL, and a maximum of 55.40 ng/mL. For instance, for a value x of 40 ng/mL, the estimate $\widehat{D}(x)$, i.e., a value higher than the mean and lower

than the third quartile, was equal to 8.22 ng/mL with 95% CI [−0.18, 16.62] and therefore statistically not significant. The estimation $\widehat{D}(x)$ for x equal to 40.5 ng/mL was statistically significant, with a value of 8.60 ng/ml and 95% CI of [0.02, 17.17]. The CATE function D quantifies the difference between the two treatment effects, i.e., after the experimental or the standard treatment. To summarize, this difference is dependent on the baseline value. Specifically, for increasing baseline values, the difference between the two treatments increases. In fact, the modeled function D is linear in baseline and the estimated $\widehat{D}(x)$ is in baseline increasing. This means that for individuals starting with higher baseline concentrations, the higher dose might be necessary when wanting to achieve a further increment (Ferrario et al. [14]). For further details, compare the freely available code at https://zenodo.org/records/6223422.

3.2. Application to a Trial in Crossover Design

The second example studies the effect of using Walkasins, which are a non-invasive neuroprosthesis, on clinical outcomes of balance and gait Koehler-McNicholas et al. [15], with freely available data at the PLOS ONE repository https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0216212.

Enrolled participants were people who experienced balance problems or were at risk of falling. Specifically, thirty-one males participated in a two-way crossover, and the participants were in a blinded, randomized study with repeated measurements at three time points in total. All participants wore Walkasins, which were turned "on" or "off" in a randomized order, with a washout period between the test days (from here, they are denoted by periods). The outcomes of interest were as follows: (i) The score of the Functional Gait Assessment (FGA). The gait speed, which was assessed under two conditions: (ii) Normal speed; (iii) Fast speed. (iv) The score of a 4-stage balance test. The working hypothesis was that participants classified as at-risk who wore Walkasins turned "on" would improve their FGA score by at least four points. Hitherto, we have shown for illustration purposes the results concerning the fast gait speed; a similar procedure can also be used for the other outcomes of interest: (i), (ii), and (iv). Participants performed a series of outcome measures while wearing the device turned OFF (Assessment #1). The assessments were repeated two additional times during the study session (Assessment #2 and Assessment #3). The first step was to import and rearrange the data (compare Listing 4). Since in the original data treatment groups were color-coded, we first rearranged the data manually, adding three columns: (i) Treatment (taking values equal to on or off); (ii) Period (taking values equal to one or two); (iii) Sequence (taking values equal to one for on-off or two for off-on). Moreover, we considered four columns with the baseline values for the four outcomes of interest (for instance, *gaitFast_base*), with the values under Assessment #1 in the original file.

Listing 4. Import and convert the data of the crossover dataset.

```
library("openxlsx")
dataDir<-"../dataDir"
data.co<-read.xlsx(file.path(dataDir,"pone.0216212.s001_pgf.xlsx"),
                   sheet="Modified")
data.co[["Treatment"]] <- as.factor(data.co[["Treatment"]])
data.co[["Period"]] <- as.factor(data.co[["Period"]])
data.co[["Sequence"]] <- as.factor(data.co[["Sequence"]])
```

We analyzed the fast gait speed as the primary outcome using a linear mixed model, taking into account the fact that the data were generated under a crossover design. Fixed effects included the treatment term, the baseline, and the adjustment for period and sequence effects. Moreover, there was a participant-specific random effect (compare Listing 5).

Listing 5. The model for the crossover dataset.

```
library("lmerTest")
fit<-lmer(gaitFast~Treatment-1+Treatment:gaitFast_base-gaitFast_base
          +Period+Sequence+(1|Subject),data=data.co)
```

The output of the calculation consisted of six estimated coefficients for the fixed effects, with 95% confidence intervals given in Table 2. While the intercept parameters were not significantly different from zero, as well as the sequence and period variables, the slope parameters were significantly different from zero, with 95% confidence intervals not crossing zero.

Table 2. Intercept and slope parameters, with 95% confidence intervals, estimated parametrically for the fast gait speed with baseline fast gait speed as a one-dimensional covariate, under a crossover design.

	Estimate	Std. Error	2.5%	97.5%
TreatmentOFF	0.05	0.14	−0.23	0.32
TreatmentON	0.06	0.14	−0.21	0.33
Period2	0.01	0.03	−0.04	0.06
Sequence2	−0.01	0.05	−0.11	0.10
TreatmentOFF:gaitFast_base	1.00	0.10	0.80	1.20
TreatmentON:gaitFast_base	1.02	0.10	0.82	1.22

Moreover, the confidence interval for the estimate $\widehat{D}(x)$ was calculated after having named the six estimated parameters from a1 to a6 (compare Listing 6 and the first column in Table 2).

Listing 6. Confidence intervals by delta method for $\widehat{D}(x)$ (second dataset).

```
coefVec <- coef(summary(fit))[, 1]
parmNames <- paste("a", 1:6, sep = "")
names(coefVec) <- parmNames
vcmat.lmm1 <- vcov(fit)
library(car)
diffFct0 <- function(baseVal)
{
  deltaMethod(coefVec, paste("a2-a1 + (a6-a5)*", baseVal), vcmat.lmm1,
              parameterNames = parmNames, level = 0.95)[, "Estimate"]
}
```

With this, the parametric estimate $\widehat{D}(x)$ was given by $(0.06 - 0.05) + (1.02 - 1) \cdot$ baseline, for any baseline value.

The baseline gait speed had a minimum value of 0.74 m/s, a first quartile of 1.18 m/s, a median of 1.4 m/s, a mean of 1.41 m/s, a third quartile of 1.58 m/s, and a maximum of 2.28 m/s. After the estimation of the fixed and the random effects, the difference function was estimated for a grid of baseline values x. As a result, the observed difference function for x equal to 0.74 m/s (i.e., for the minimum of x) was not statistically significant, with an estimated value of 0.03 m/s and 95% CI of [−0.09, 0.14]. The observed difference function for x equal to 1.41 m/s (i.e., for the mean) was not statistically significant, with an estimated value of 0.04 m/s and 95% CI of [−0.01, 0.09]. Finally, the observed difference function for x equal to the maximum baseline value of 2.28 m/s was likewise not statistically significant, with an estimated value of 0.05 m/s and 95% CI of [−0.12, 0.23]. Together, these results imply that the values of the baseline gait speed are not crucial in predicting whether the difference between the two treatment effects (on minus off) is positive. The generated code for this analysis is available online on zenodo.org: https://zenodo.org/records/13945382.

3.3. Simulation Study

In order to evaluate the performance of the described method in finite-sample settings, we conducted a simulation study. Specifically, the goal of the simulation was to evaluate the bias of the parametric estimator for increasing sample size, n, and under different models in the setup analogy to the dataset in Section 3.1.

For this, we generate data from models with known parameters. Specifically, we defined two lists, which contained a set of values for the intercept and for the slope parameters. These lists represented different scenarios or conditions we wanted to simulate. Subsequently, some parameters were set: the number of generated datasets equal to 1000 and four different possible sample sizes, which were equal to 60, 120, 180, and 10,000. Together, many datasets, depending on the different parameters, were generated. Moreover, we considered two treatment groups and two time points. Half of the individuals were assigned to the standard treatment and half to the experimental treatment. A one-dimensional covariate x was considered, which was generated by a uniform distribution between 0 and 40. Random errors (epsilon) and random intercepts were generated as independent and standard normally distributed. The response variable y was drawn based on the linear mixed-effects model. The models included fixed effects for the covariate x for a new categorical variable, which encodes the combinations of the two treatments and two time points, as well as a random intercept for the individuals. This resulted in four intercept parameters and four slope parameters for the fixed effects. For the given fit, we extracted the estimated coefficients. Finally, the bias between the true coefficients and the estimated coefficients for each generated dataset was calculated, which was given as a percentage. A seed was set for reproducibility, the mapply R function was used to apply the function simulating the data to different sets of true intercept and slope parameters (R version 4.4.1).

For the first scenario, the four intercept parameters were $\beta_0 = (10, 12, 13, 11)^T$, and the four slope parameters were $\beta_1 = (0.3, 0.6, 0.2, 0.6)^T$. For the second scenario, the four intercept parameters were $\beta_0 = (9, 12, 14, 10)^T$, and the four slope parameters were $\beta_1 = (0.4, 0.5, 0.3, 0.7)^T$. For the last scenario, the four intercept parameters were $\beta_0 = (9.5, 11.5, 12.5, 10.5)^T$, and the four slope parameters were $\beta_1 = (0.5, 0.6, 0.3, 0.8)^T$. The results of this simulation are summarized in two tables (Tables 3 and 4), which reports the empirical bias and the empirical standard deviation in percentage for the estimated slope and intercept coefficients under the three working models.

Table 3. Estimation of average bias with standard deviation for the estimated intercept parameters of the given true coefficients. The estimated parameters b0 are given for the first time point and for the second time point under the two treatments. All the estimated numbers are given in percentages. The number of generated datasets was equal to 1000.

ss *	b0 t1 (0) True Coef	b0 t1 (0) Bias m	b0 t1 (0) Bias sd	b0 t2 (0) True Coef	b0 t2 (0) Bias m	b0 t2 (0) Bias sd	b0 t1 (1) True Coef	b0 t1 (1) Bias m	b0 t1 (1) Bias sd	b0 t2 (1) True Coef	b0 t2 (1) Bias m	b0 t2 (1) Bias sd
60	10.00	−0.11	3.86	12.00	0.03	3.25	13.00	−0.10	2.77	11.00	−0.04	3.50
120	10.00	−0.08	2.52	12.00	0.01	2.21	13.00	0.00	1.99	11.00	0.05	2.38
180	10.00	0.02	2.17	12.00	0.03	1.77	13.00	0.08	1.61	11.00	0.05	1.89
10,000	10.00	0.01	0.29	12.00	−0.00	0.25	13.00	0.01	0.21	11.00	−0.00	0.25
60	9.00	−0.31	4.29	12.00	−0.02	3.14	14.00	−0.05	2.68	10.00	−0.10	3.78
120	9.00	−0.14	2.91	12.00	−0.11	2.15	14.00	0.00	1.80	10.00	−0.05	2.47
180	9.00	0.09	2.33	12.00	0.05	1.69	14.00	−0.06	1.57	10.00	−0.12	2.15
10,000	9.00	−0.02	0.31	12.00	−0.01	0.23	14.00	−0.01	0.20	10.00	−0.01	0.29
60	9.50	−0.03	3.95	11.50	0.02	3.16	12.50	−0.08	2.94	10.50	−0.06	3.39
120	9.50	−0.01	2.74	11.50	0.01	2.24	12.50	0.02	2.05	10.50	0.03	2.40
180	9.50	−0.05	2.22	11.50	0.00	1.80	12.50	−0.02	1.71	10.50	−0.08	2.01
10,000	9.50	−0.01	0.30	11.50	0.00	0.24	12.50	−0.00	0.23	10.50	0.02	0.27

* ss: sample size. true coef: true coefficients. bias m: estimated average of the bias. bias sd: estimated standard deviation of the bias. b0: intercept parameters. t1: first time point. t2: second time point. (0): treatment (0). (1): treatment (1).

Table 4. Estimation of average bias with standard deviation for the estimated slope parameters of the given true coefficients. The estimated parameters b1 are given for the first time point and for the second time point under the two treatments. All the estimated numbers are given in percentages. The number of generated datasets was equal to 1000.

ss *	b1 t1 (0) True Coef	b1 t1 (0) Bias m	b1 t1 (0) Bias sd	b1 t2 (0) True Coef	b1 t2 (0) Bias m	b1 t2 (0) Bias sd	b1 t1 (0) True Coef	b1 t1 (1) Bias m	b1 t1 (1) Bias sd	b1 t2 (1) True Coef	b1 t2 (1) Bias m	b1 t2 (1) Bias sd
60	0.30	0.11	5.55	0.60	−0.05	2.79	0.20	0.21	7.78	0.60	0.02	2.80
120	0.30	0.16	3.59	0.60	0.01	1.90	0.20	−0.11	5.69	0.60	-0.05	1.89
180	0.30	−0.08	3.17	0.60	−0.02	1.56	0.20	−0.18	4.47	0.60	−0.07	1.50
10,000	0.30	−0.02	0.40	0.60	0.00	0.21	0.20	−0.01	0.59	0.60	0.00	0.20
60	0.40	0.29	4.15	0.50	−0.04	3.25	0.30	0.02	5.39	0.70	0.09	2.31
120	0.40	0.11	2.86	0.50	0.03	2.25	0.30	0.00	3.66	0.70	0.03	1.56
180	0.40	−0.10	2.31	0.50	−0.02	1.79	0.30	0.04	3.13	0.70	0.04	1.29
10,000	0.40	0.02	0.30	0.50	0.01	0.24	0.30	0.02	0.40	0.70	0.01	0.18
60	0.50	−0.01	3.19	0.60	−0.00	2.61	0.30	0.14	5.33	0.80	0.02	1.95
120	0.50	−0.00	2.27	0.60	0.01	1.84	0.30	0.05	3.72	0.80	−0.03	1.39
180	0.50	−0.00	1.83	0.60	−0.04	1.53	0.30	−0.01	3.07	0.80	0.03	1.16
10,000	0.50	0.01	0.24	0.60	0.00	0.20	0.30	−0.01	0.42	0.80	−0.01	0.16

* ss: sample size. true coef: true coefficients. bias m: estimated average of the bias. bias sd: estimated standard deviation of the bias. b1: slope parameters. t1: first time point. t2: second time point. (0): treatment (0). (1): treatment (1).

By the maximium likelihood theory the used estimates of the model parameters are consistent and asymptotically normal distributed with the regular rate of convergence $\sqrt{n}$, hence the bias and the standard deviation of the bias should converge to zero. We see that the bias is still unstable for small sample sizes (60-120-180), varying around a value close to zero. For instance, for a true intercept parameter equal to 13 the estimated bias was equal to -0.10, -0.00 and 0.08 for n equal to 60, 120 or 180, respectively. Moreover, as expected by the theory, the bias is getting smaller when the sample size is sharply increased to 10,000. Here, the bias for the intercept parameter was equal to 0.01.

We also see that the estimated standard deviation is high for small sample sizes: for instance, for a true slope parameter equal to 0.2, the standard deviation of the bias is equal to 7.78, 5.69, 4.47 for n equal to 60, 120, or 180. For a sample size of 10,000 the standard deviation is equal to 0.59, which decreases by increasing sample size, as expected because of the strong law of large numbers.

Analogously, for a true slope parameter equal to 0.6, the estimated bias was oscillating to the values -0.02, -0.05, -0.07 for n equal to 60, 120, or 180, respectively, while the bias for a sample size of 10,000 decreased to -0.00. This is also expected by the theory, specifically by the consistency of the estimator.

The entire generated code for these simulations is available online at https://zenodo.org/records/14725244.

4. Discussion

In this article, we estimated the conditional average treatment effect function (CATE function D) involving a one-dimensional covariate, and we illustrate this procedure by real data examples through a re-evaluation of publicly available data. Specifically, we worked on a dataset in the context of a supplementation by vitamin D. Here, the participants receive standard treatments and not personalized treatment. However, by the estimation of individualized responses to the classical treatments we have new criteria for a future personalized treatment. Moreover, the second real data example concerning the Walkasins represents a possible application of a potential personalized medicine. We, therefore, believe that the described procedures can be useful in many disciplines of applied statistics.

It is important to note that the choice of adequate individual baseline covariates should be based on the clinical relevance of the pre-treatment information and not be chosen in a data-driven manner. Moreover, we want to underline that a baseline covariate could be strongly associated with one outcome of interest (and therefore be known for this in the

literature), but at the same time, this baseline covariate may not be a valuable predictor of the outcome of interest.

To estimate an ITR, it is not necessary, in general, to use an estimate of the outcome Y. However, the estimated treatment success can be valuable per se. The problem of estimating an ITR is related to a classification problem. However, the underlying data might be different since the true class memberships of the training data, i.e., the true better individual treatments, are not known. In the example of supplementation by vitamin D the training data are parts of a parallel design study, so that only the outcome of one randomly assigned treatment per participant can be used. The second example concerning the Walkasins was based on a crossover study, so that for each participant the better treatment choice is only available with respect to a negligible carry over effect.

The ITRs, as discussed here, involved two possible treatments: A standard and an experimental treatment; however, further scenarios could be realistic in case of a complex design also involving three or more treatments. Also, in such a scenario under a different, more complex design, estimation of the CATE function D could be of interest and should be investigated accordingly. Moreover, an unbalanced design with respect to the treatments could represent a further issue, ideally to be examined in advance.

Generally speaking, the delta method uses a Taylor expansion yielding point estimate and confidence intervals for a (non-linear) transformation (here, for the CATE function D) of a random variable (here, $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$, the parameter estimators, for instance) [16]. This works when the random variable is asymptotically normal. Other alternatives could be possible in order to construct confidence intervals. A classical alternative would be the construction of the confidence intervals based on a (multivariate) t-distribution. In the context of mixed models, the estimation of the degree of freedom of the t-distribution could become challenging. Another approach would be bootstrapping. However, the delta method typically requires fewer computation and perform well with small sample sizes (which is the case here).

Finally, there are further possible approaches for estimation of the CATE function based on so-called meta-algorithms that have been proposed in the literature (Caron et al. [10]). These meta-algorithms are referred to as Meta-Learners and are machine learning algorithms. Some examples of these are random forests, LASSO regression, neural networks, gradient boosting trees, etc.

In view of the results of the simulation study, and also supported by theoretical considerations, we see that the sample size should not be too small, and we therefore recommend investing time in calculating a proper sample size to better achieve asymptotic results. The simulation study focused on the bias of the proposed estimator. This is a start for further simulations, which can be easily conducted since we share the entire code.

A natural extension of the discussed procedures would be to include a two-dimensional or a higher dimensional covariate vector for exploring how treatment effects vary depending on a pre-treatment covariate vector. However, dealing with a high dimensional vector is a non-trivial step for many reasons. There is, firstly, the issue of how to model the relationship between the outcome of interest and the covariate vector. In fact, the CATE function D could take a complex form in a high dimensional context. Then, overfitting issues could arise if the covariate vector dimension is not small. Moreover, these methods are subject to the curse of dimensionality, requiring the recruitment of a much larger number of individuals to draw conclusions. A possible correlation between the components of the covariate vector poses challenges when making inferences, or at least such correlations could be taken into account for appropriate multiplicity corrections (i.e., less conservative than, for instance, by Bonferroni corrections). Here, the multiple marginal model approach could be helpful [17].

Theoretical and applied procedures for considering a higher dimensional covariate vector warrant further research.

Author Contributions: P.G.F.: Data search, implementation and data analysis/interpretation, simulation, and writing the manuscript. M.D.: Implementation, simulation, comments on the code and on the manuscript. C.R.: Implementation and results interpretation, supervision, comments on the code and on the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in: https://zenodo.org/records/13945382 (accessed on 17 October 2024); https://zenodo.org/records/14725244 (accessed on 23 January 2025); https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117123 (accessed on 17 October 2024); https://zenodo.org/records/6223422 (accessed on 17 October 2024); https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0216212 (accessed on 17 October 2024).

Conflicts of Interest: The authors declare no potential conflicts of interests.

Appendix A

Using the package `plyr`, we rearrange the data of the first real data example concerning vitamin D as follows.

Listing A1. Rearrange the data of the vitamin D dataset.

```
dataDir<-"../dataDir"
data.vitd.3<-read.csv2(file.path(dataDir,
                                 "journal.pone.0117123.s006.csv"))

# from broad to long form
library(plyr)
plyFct <- function(dataSet){ds2 <- dataSet[-1, ];
ds2[["baseline"]] <- dataSet[1, "Serum.25.OH.D.ng.mL"]; ds2}
data.vitd.3nobase <- ddply(data.vitd.3, .(ID), plyFct)
write.csv2(data.vitd.3nobase,file.path(dataDir,
                                       "journal.pone.0117123.s006_pgf.csv"),
           quote=FALSE,row.names =FALSE )
```

References

1. Cornelis, M.C.; El-Sohemy, A.; Campos, H. Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. *Am. J. Clin. Nutr.* **2007**, *86*, 240–244.
2. Berry, S.E.; Valdes, A.M.; Drew, D.A.; Asnicar, F.; Mazidi, M.; Wolf, J.; Capdevila, J.; Hadjigeorgiou, G.; Davies, R.; Al Khatib, H.; et al. Human postprandial responses to food and potential for precision nutrition. *Nat. Med.* **2020**, *26*, 964–973.
3. Ritz, C.; Astrup, A.; Larsen, T.M.; Hjorth, M.F. Weight loss at your fingertips: Personalized nutrition with fasting glucose and insulin using a novel statistical approach. *Eur. J. Clin. Nutr.* **2019**, *73*, 1529–1535.
4. Zeisel, S.H. Precision (Personalized) Nutrition: Understanding Metabolic Heterogeneity. *Annu. Rev. Food Sci. Technol.* **2020**, *11*, 71–92.
5. Athey, S.; Imbens, G. Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7353–7360. https://doi.org/10.1073/pnas.1510489113.
6. Matsouaka, R.A.; Li, J.; Cai, T. Evaluating marker-guided treatment selection strategies. *Biometrics* **2014**, *70*, 489–499. https://doi.org/10.1111/biom.12179.
7. Künzel, S.R.; Sekhon, J.S.; Bickel, P.J.; Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4156–4165. https://doi.org/10.1073/pnas.1804597116.

8. Cai, T.; Tony Cai, T.; Guo, Z. Optimal statistical inference for individualized treatment effects in high-dimensional models. *J. R. Stat. Soc. Ser. B* **2021**, *83*, 669–719. https://doi.org/10.1111/rssb.12426.
9. Du, Y.; Chen, H.; Varadhan, R. Lasso estimation of hierarchical interactions for analyzing heterogeneity of treatment effect. *Stat. Med.* **2021**, *40*, 5417–5433. https://doi.org/10.1002/sim.9132.
10. Caron, A.; Baio, G.; Manolopoulou, I. Estimating Individual Treatment Effects using Non-Parametric Regression Models: A Review. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2022**, *185*, 1115–1149. https://doi.org/10.1111/rssa.12824.
11. Rosenbaum, P.R.; Rubin, D.B. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **1983**, *70*, 41–55.
12. Steenhoff, A.P.; Schall, J.I.; Samuel, J.; Seme, B.; Marape, M.; Ratshaa, B.; Goercke, I.; Tolle, M.; Nnyepi, M.S.; Mazhani, L.; et al. Vitamin D3 supplementation in Batswana children and adults with HIV: A pilot double blind randomized controlled trial. *PLoS ONE* **2015**, *10*, e0117123.
13. Royston, P.; Sauerbrei, W. Interaction of treatment with a continuous variable: Simulation study of power for several methods of analysis. *Stat. Med.* **2014**, *27*, 4695–4708. https://doi.org/10.1002/sim.6308.
14. Ferrario, P.G.; Watzl, B.; Ritz, C. The role of baseline serum 25(OH)D concentration for a potential personalized vitamin D supplementation. *Eur. J. Clin. Nutr.* **2022**, *76*, 1624–1629.
15. Koehler-McNicholas, S.R.; Danzl, L.; Cataldo, A.Y.; Oddsson, L.I.E. Neuromodulation to improve gait and balance function using a sensory neuroprosthesis in people who report insensate feet A randomized control cross-over study. *PLoS ONE* **2019**, *14*, e0216212
16. Ver Hoef, J.M. Who Invented the Delta Method? *Am. Stat.* **2012**, *66*, 124–127.
17. Hothorn, T.; Bretz, F.; Westfall, P. Simultaneous inference in general parametric models. *Biom. J./Biom. Z.* **2008**, *50*, 346–363. https://doi.org/10.1002/bimj.200810425.