Evaluation of Adherence to Nutritional Intervention Through Trajectory Analysis

B. Sevilla-Villanueva, K. Gibert, M. Sànchez-Marrè, M. Fitó, and M. I. Covas

Abstract—Classical pre–post intervention studies are often analyzed using traditional statistics. Nevertheless, the nutritional interventions have small effects on the metabolism and traditional statistics are not enough to detect these subtle nutrient effects. Generally, this kind of studies assumes that the participants are adhered to the assigned dietary intervention and directly analyzes its effects over the target parameters. Thus, the evaluation of adherence is generally omitted. Although, sometimes, participants do not effectively adhere to the assigned dietary guidelines. For this reason, the trajectory map is proposed as a visual tool where dietary patterns of individuals can be followed during the intervention and can also be related with nutritional prescriptions. The trajectory analysis is also proposed allowing both analysis: 1) adherence to the intervention and 2) intervention effects. The analysis is made by projecting the differences of the target parameters over the resulting trajectories between states of different time-stamps which might be considered either individually or by groups. The proposal has been applied over a real nutritional study showing that some individuals adhere better than others and some individuals of the control group modify their habits during the intervention. In addition, the intervention effects are different depending on the type of individuals, even some subgroups have opposite response to the same intervention.

Index Terms—Adherence analysis, clustering, interpretation, pre–post studies, trajectory analysis.

I. INTRODUCTION

THE paradigm of health science is promoting the personalized medicine where the nutritional habits become an important factor. In order to understand *why similar diets have different effect on apparently similar people*, several intervention studies have been performed. An intervention study answers *"What is the effect of intervention?"* by examining whether participants change during the course of the study, and the possible improvements or worsenings are attributed to the intervention [1].

Data included in such studies come from surveys and clinical tests. In most cases, an important number of characteristics are measured. These attributes have different origins, such as anthropometric, socio-demographic, habits, and clinical blood, urine or genetic tests. The analysis of these studies is complex due to many reasons. Among them, the relatively small effect of nutrients on physiological parameters; effects can be reflected in multiple biomarkers; and also, complex interactions between nutrients or other environmental factors and genes which are difficult to detect [2].

The classical statistical approach is to consider only numerical attributes, compute differences and search for a model relating the diet with those differences. Although these studies are usually for drug development, health research is currently interested in getting scientific evidence from a wider spectrum of health related issues like neurorehabilitation, nutrition, mental health, behavior, or social science [3]. For these situations, the traditional techniques do not always extract all relevant information from data [4].

Nowadays, it is well known that Knowledge discovery provides a good framework to complex phenomena [4], when contextual factors that can interact with the intervention itself are not all well known and not easy to measure or too much to be explicitly considered. Some references are found about studies which introduce some artificial intelligence (AI) techniques. These works state that the use of AI can be useful to extract knowledge from this kind of data and understand the risk of a specific disease. In a study about hypertension [5], the results obtained by a logistic regression and two decision trees were compared. In [6], statistical models and association rules were used to study the comorbidity of attention deficit hyperactivity disorder (ADHD).

With the Humane Genome Project in 90's, there is an increasing interest of extract knowledge from genetic data. Most works analyze microarray data directly. Basic hypothesis test are often used to test significant changes in gene expression [7].

Since our data comes from a Mediterranean diet with olive oil intervention study, the following works from important trials are included about the benefits of this intervention. The

B. Sevilla-Villanueva and K. Gibert are with the Knowledge Engineering and Machine Learning Group (KEMLG), Department of Statistics and Operations Research (EIO), Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Barcelona 08034, Spain (e-mail: bea.sevilla@gmail.com; karina@upc.edu).

M. Sànchez-Marrè is with the Knowledge Engineering and Machine Learning Group, Department of Computer Science (CS), Universitat Politècnica de Catalunya - BarcelonaTech, Barcelona 08034, Spain (e-mail: miquel@cs.upc.edu).

M. Fitó and M. I. Covas are with the Hospital del Mar Institute of Medical Research Investigations (IMIM), Barcelona 08003, Spain, and also with the CIBER of the Obesity Physiopathology and Nutrition (CIBEROBN), Madrid 28029, Spain (e-mail: mfito@imim.es; MCovas@imim.es).

European project EUROLIVE carried out several studies with the objective to obtain scientific evidence on the impact of olive oil, and its phenolic compounds, on oxidative stress and damage in several European populations [8]. These works were analyzed with classical statistics. PREDIMED project is a long-term nutritional intervention study. It aimed to assess the efficacy of Mediterranean diet in the prevention of cardiovascular diseases. Also, these studies were analyzed with classical statistics (one-factor analysis, Kruskal-Wallis test or χ^2-Independence) like in [9].

Within the specific field of nutrigenomics, only two references have been found were AI is used in a pre–post analysis with humans. The first study is about the effect of different low-energy diets on gene expression [10]. A hierarchical clustering was performed on genes showing changes in their expression during the dietary intervention and which of them changed in similar way. The second study [11] focuses on clustering gene expression obtained during low-calorie diet intervention to aid in the prediction of 6-month weight loss maintenance. They clustered the patients according to their gene expression using an hierarchical clustering with average linkage. This study points to a need of *local studies over the different groups of intervention*.

Usually, the adherence to the intervention is assumed in the analysis of these studies. Nevertheless, a complete adherence could not occur if the participants do not follow all the dietary guidelines. Our proposal is to evaluate the adherence for a further analysis of the intervention effects with more precision. Concretely, a previous classification of participants is proposed to locally analyze both the adherence to the intervention and its effects on each type of individual [12]. For this, the *Trajectory Analysis* is introduced to analyze how the participants have evolved during the intervention. This analysis characterizes the resulting trajectories from the previous build *Trajectory Map*.

The present proposal is applied on a real pre–post study. Even though the methodology is conceived for pre–post intervention studies, this can be used with the desired temporal granularity to analyze how the diet of the participants are evolving. This proposal can be included in a monitoring tool to analyze the diet evolution if it receives the participants' meals as an input. The resulting information can be valuable for both participants and health professionals. Participants can observe how their habits have changed and correct their behavior if they are getting away from their prescribed intervention, giving way to a self health-management tool. Health professionals can follow the evolution of a participant along the time including also other information such as health status, blood and urine biomarkers or gene expression and preferences. Even, this tool can be linked with a diet recommender as the one proposed in the project Diet4You [13].

The structure of this paper is the following: the proposed methodology is presented in Section II; the case study and the results are given in Section III with the consequent discussion in Section IV and finally, conclusions and future work are in Section V.
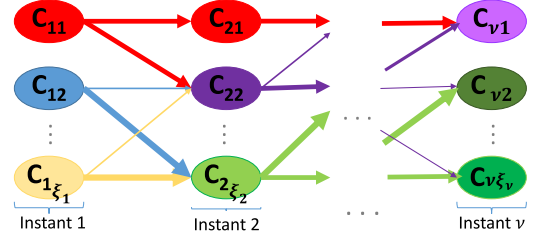

Fig. 1. General Schema of the trajectory map: Each column represents the states ($P_t = \{c_{t1}, \ldots, c_{t\xi_t}\}$) at certain instant t.

II. METHODOLOGY

Lets assume that the set of individuals $\mathcal{I}$ are initially classified in terms of their situations (good health conditions, obesity problems, etc.) both at the beginning and at the end of the intervention. Let us call *state* each of this initial and/or final situations (healthy state, unhealthy state, etc.). The group of individuals sharing a certain state at a given instant t is the extension of a class defining the state. The set of possible states of individuals at a given timestamp represents a partition of $\mathcal{I}$ that can be created using knowledge-based, data-driven techniques, ontology-driven, etc. In fact, there can be as many sets of states as considered timestamps, which can keep equal along time or not.

A local analysis of the changes during the intervention is performed (*Trajectory Analysis*) considering how individuals change state along time and the assigned intervention. First, we propose to build a Trajectory Map that graphically represents all possible states at each timestamp. From this representation, it is possible to analyze how the individuals in one state are distributed in the states of the next timestamp and so on. One can observe how the intervention modifies the states of the individuals by the intensional description of these classes. Fig. 1 represents the general schema of the Trajectory Map. There are ν timestamps that can have different states or classes ($P_t = \{C_{t1}, \ldots, C_{t\xi_t}\}, t = 1 : \nu$). The width of edges represents the number of individuals that moves from one state to another between two instants. Whereas the states represents a situation related with some latent attribute (healthy life style, or nutritional situation, etc.), colors can be associated to each state accordingly.

In the case of two timestamps, the trajectory map can be graphically represented with states previous to intervention at the left-hand side of the graph, posterior states at the right-hand side, and corresponding edges connecting one previous state with one posterior state.

On top of this representation, both the adherence to the intervention and their effects can be evaluated with the proposed *Table of Profile Comparisons* of Section II-B.

The proposed methodology is: Given,
$\mathcal{I} = \{i_1, \ldots, i_n\}$ the set of individuals. $\mathcal{X}$ a set of attributes describing $\mathcal{I}$. $\mathcal{P}_o$ a partition of $\mathcal{I}$ representing the set of previous states. $\mathcal{P}_f$ a partition of $\mathcal{I}$ representing the set of posterior states. T the assigned intervention to each individual,

1) Determine δ as the minimum number of individuals per trajectory to be retained.

2) Build Ψ the set of characteristic trajectories:
$\Psi = \{\phi_1, \ldots, \phi_{\xi_\Psi}\}; \Psi \subseteq \mathcal{P}_o \times T \times \mathcal{P}_f$;
$\phi \in \Psi$ iff $n_\phi \geq \delta$ where $n_\phi = \text{card}(\phi)$.
3) Build the trajectory map by representing $\mathcal{P}_o \times T$ on the left hand side and $\mathcal{P}_f$ on the right. Represent each trajectory $\phi \in \Psi$ with an edge between the involved states of $\mathcal{P}_o \times T$ and $\mathcal{P}_f$ with a width proportional to n_ϕ.
4) Analysis of adherence to the intervention (T)
 a) Identify $\mathcal{X}^T \subset \mathcal{X}$, the subset of attributes directly related with the intervention (food consumption, drugs in pharmacological treatments, etc.).
 b) Analysis of the adherence to the intervention T over $\mathcal{P}_o \times \mathcal{P}_f$. Synthesize the relationships among $\mathcal{P}_o, \mathcal{P}_f$ and T in terms of:
  i) Interpretations $\Upsilon_{\Psi, \mathcal{X}_o^T}$ and $\Upsilon_{\Psi, \mathcal{X}_f^T}$ of Ψ by applying NCI-IMS methodology (see Section II-A) to the measures of $\mathcal{X}^T$ taken before and after the intervention.
  ii) Changes observed in the *table of profile comparisons* $f_{D,T}$ between $\Upsilon_{\Psi, \mathcal{X}_o^T}$ versus $\Upsilon_{\Psi, \mathcal{X}_f^T}$ (see Section II-B).
  iii) Evaluation of the adherence to the intervention associated to the trajectories.
(5) Analysis of the Intervention effects characterizing each trajectory in Ψ:
 a) Identify $\mathcal{X}^E \subset \mathcal{X}$, the subset of attributes showing the effects of the intervention (biomarkers, gene expression, etc.)
 b) Interpret Ψ by applying NCI-IMS (Section II-A) to the measures of $\mathcal{X}^E$ taken before and after the intervention : Get $\Upsilon_{\Psi, \mathcal{X}_o^E}$ and $\Upsilon_{\Psi, \mathcal{X}_f^E}$.
 c) Build the table of profile comparisons $f_{D,E}$ between $\Upsilon_{\Psi, \mathcal{X}_o^E}$ versus $\Upsilon_{\Psi, \mathcal{X}_f^E}$.
 d) Synthesize the changes of $f_{D,E}$ in terms of intervention effects considering the adherence.

Thus, the resulting profiles at the beginning and at the end of the study are compared depending on the assigned intervention. This allows detecting the degree of adherence using the related attributes to the intervention (in this case, the diet habits). Also, the intervention effects can be evaluated using other attributes, such as the biomarkers or gene expression, when adhesion is maintained. Finally, to evaluate the benefits of this proposal the trajectory analysis is compared with the analysis of changes only considering the global intervention groups.

A. Interpretation Methodology

The interpretation methodology NCI-IMS [12], [17] is used to interpret the classes of a partition P in terms of a set of attributes $\mathcal{X}$. NCI-IMS gives an interpretation $\Upsilon_{P,\mathcal{X}}$ containing the significant attributes in each class and indicating the *sense* of their significance: significant higher ($\uparrow$) or lower ($\downarrow$) values than the sample or non-significance ($-$).

B. Table of Profile Comparisons

Given $\Upsilon_{P_o,\mathcal{X}_o}$ and $\Upsilon_{P_f,\mathcal{X}_f}$, the interpretations of partitions P_o and P_f in term of $\mathcal{X}_o, \mathcal{X}_f$ respectively,

	Sense of X at instant t		
	$\downarrow$	$-$	$\uparrow$
$\downarrow$	$-$	$\uparrow$	$\uparrow$
$-$	$\downarrow$	$-$	$\uparrow$
$\uparrow$	$\downarrow$	$\downarrow$	$-$

(Sense of X at instant t-1)

$\uparrow$ t significant higher than t-1.
$\downarrow$: t significant lower than t-1.
$-$:No significant difference between t-1 and t.

1) Use Table I to build a table of profile comparisons f_D, a vector indicating significant increase or decrease of every attribute after the intervention.
2) The pre–post changes can be measured as the number of attributes with significant increasing and the number of attributes with significant decreasing.
3) Whereas additional expert knowledge about semantics of attributes (healthy/unhealthy, prescribed/prohibit, etc.) is introduced, new indicators on number of total changes per group of attributes can be computed. According to the attributes used, different information can be extracted from the f_D, for example:
 a) When attributes describe the intervention and are qualified as prescribed or prohibited items (healthy/unhealthy food), the *adherence* to the intervention can be measured as the number of prescribed items significantly increased plus the number of discouraged items decreased.
 b) When attributes are biomarkers and qualified as beneficial or damaging, the *effect* of the items can be measured as the number of beneficial biomarkers significantly increased plus the number of damaging biomarkers significantly decreased.

III. APPLICATION OF THE TRAJECTORY ANALYSIS TO A NUTRITIONAL CASE STUDY

A. Case Study Description

The case study goal was to analyze the effect of the Mediterranean diet and virgin olive oil on healthy people and, particularly, to assess whether their benefits could be mediated through changes in the expression of atherosclerosis-related genes as it presented in [9], [14]. For this study, a randomized, parallel, controlled clinical trial with the following three dietary interventions was performed:
1) VOO: Mediterranean diet + Virgin olive oil.
2) WOO: Mediterranean diet + Washed olive oil.
3) Ctrl (Control group): Follow the usual diet.
First, an inclusion test defined the individuals that can be recruited. After, the individuals were selected and randomly split into the intervention groups (89 healthy individuals aged 20–50 years). Then, there were two visits: before and after the dietary intervention. The data comprised biomarkers of blood, urine, and gene expression obtained by clinical test (gene expression only available in 47 individuals) and self-reported surveys asking for socio-demographic data, health and familiar history, drug treatments, tobacco, alcohol, dietary, and exercise habits. The dietary habits correspond to the self-reported p14-item questionnaire published in [15] (see Table II). Authors are aware

TABLE II
MEDITERRANEAN DIET INDICATORS (P-14 SURVEY)

Code	Name	Description	Value	Semantic	Med. index
p14_1	MainOliveOil	Use of olive oil as main fat	Yes/No	Maximize	Yes
p14_2	OliveOil	4 or more spoons/day of olive oil	Yes/No	Maximize	Yes
p14_3	Vegetables	2 or more pieces/day of vegetables	Yes/No	Maximize	Yes
p14_4	Fruit	3 or more pieces/day of fruit (including natural juices)	Yes/No	Maximize	Yes
p14_5	RedMeat	1 or more portions/day of red meat, hamburgers or sausages	Yes/No	Minimize	No
p14_6	Butter	1 or more portions/day of butter, margarine or cream	Yes/No	Minimize	No
p14_7	GasDrinks	1 or more glasses/day of gas drinks and/or sugary drinks	Yes/No	Minimize	No
p14_8	Wine	7 or more glasses/week of wine	Yes/No	Maximize	Yes
p14_9	Legumes	3 or more portions/week of legumes	Yes/No	Maximize	Yes
p14_10	Fish	3 or more portions/week of fish, shellfish or sea food	Yes/No	Maximize	Yes
p14_11	Com.Bakery	2 or more portions/week of commercial bakery	Yes/No	Minimize	No
p14_12	Nuts	3 or more portions/week of nuts	Yes/No	Maximize	Yes
p14_13	WhiteMeat	Consume preferable of white meat (chicken, turkey, rabbit) instead of red meat, pork, hamburgers or sausages	Yes/No	Maximize	Yes
p14_14	Sauce	2 or more portions/week of homemade tomato sauce	Yes/No	Maximize	Yes

of the limitations of the self-reported data [16]. A more robust solution should be to have smart devices and monitoring sensors.

The dataset contained 242 attributes. In preprocessing, 141 attributes not providing useful information for clustering were cleaned (51 empty, 34 constant, 31 redundant, 25 irrelevant this research target, i.e., address). The 65 remaining attributes basically referred to two semantic aspects:

1) Baseline block: it describes the health condition (biometric, tobacco, socio-demographic, diseases, drugs, and biomarkers).
2) Habits block: food habits and physical activity.

According to the integrative multiview clustering [17], individuals were clustered twice, one per block. Ward's method [18] with the Gower's dissimilarity coefficient [19] were used. The resulting partitions were crossed in order to obtain an integrated partition. The resulting partition $\mathcal{P}_o$ defines the initial states of individuals at the beginning of the study and it is composed by eight classes: M-WMbased, M-WMwSugars, YW-UH, YW-WMbased, YW-WMwSugars, YW-UH, WM-WMbased, and WM-WMwSugars. The process was repeated with the measures at the end of the intervention and the resulting partition $\mathcal{P}_f$ defines the final states composed by 12 classes: M-H↑OO, M-HB, M-BasicBak, M-ProtoCaloric, W-H↑OO, W-HB, W-BasicBak, W-ProtoCaloric, WM-H↑OO, WM-HB, WM-BasicBak, and WM-ProtoCaloric.

B. Pre–Post Trajectory Analysis

In this section, the initial state (P_o) is compared against the final state of the individuals (P_f) taking into account the intervention (T) using the trajectory analysis. The Trajectory Map (see Fig. 2) is build only with the trajectories having more than one individual ($\delta = 2$).

In this work, the results of our approach (local to each trajectory) are compared with the ones obtained under direct use of the global intervention groups. In order to show the differences in each trajectory, the profiles of each trajectory (before and after the intervention) are assessed. These profiles are made using the NCI-IMS methodology, the differences between the profiles before and after the intervention are assessed. Table III collects

the results: top of table (first three rows) contains the comparison of the global intervention groups; bottom of table contains the comparison of profiles of each trajectory. The first 14 attributes (columns) correspond to dietary habits for analyzing the adherence. Next columns are counters (number of attributes that have increased and decrease the consumption; increases of healthy items and decreases of unhealthy items). On the right part there are six blood biomarkers, one urine biomarker, and finally, the comparison of the gene expression of five genes.

IV. DISCUSSION

The objective of this *Trajectory Analysis* is to observe how the individuals have evolved during the intervention. Since the intervention is taken into account, whether individuals of one class behaves differently depending on the assigned intervention (T) can be analyzed.

The trajectory map (see Fig. 2) shows how the individuals belong to one or other final class depending on the initial state and the assigned intervention. First, from this figure, it is possible to observe that both interventions and control are not balanced inside each class. Broadly, men groups are the most balanced ones (33% VOO, 29% WOO, 37% Ctrl), Young women groups (31.4% VOO, 43.1% WOO and 25.5% Ctrl), and Women with menopause (43% VOO, 7% WOO, 50% Ctrl). Comparing the subclasses, YW-UH contains the greatest proportion of WOO while WM-WMwSugars contains the smallest. Second, in general, individuals that are assigned to an intervention are more divided in the final classes than the individuals assigned to the control group which are more concentrated in some final class indicating that the $Ctrl$ groups are more stable.

A. Adherence to the Intervention

Diet profiles are compared analyzing the changes of diet patterns regarding those items that are more or less consumed (first 14 columns of Table III). Since these diet items are directly related with the intervention, this analysis leads to evaluate the adherence by observing changes in diet indicators after with respect to before the intervention. This analysis is enriched with the information of healthy and unhealthy items (see Table II).
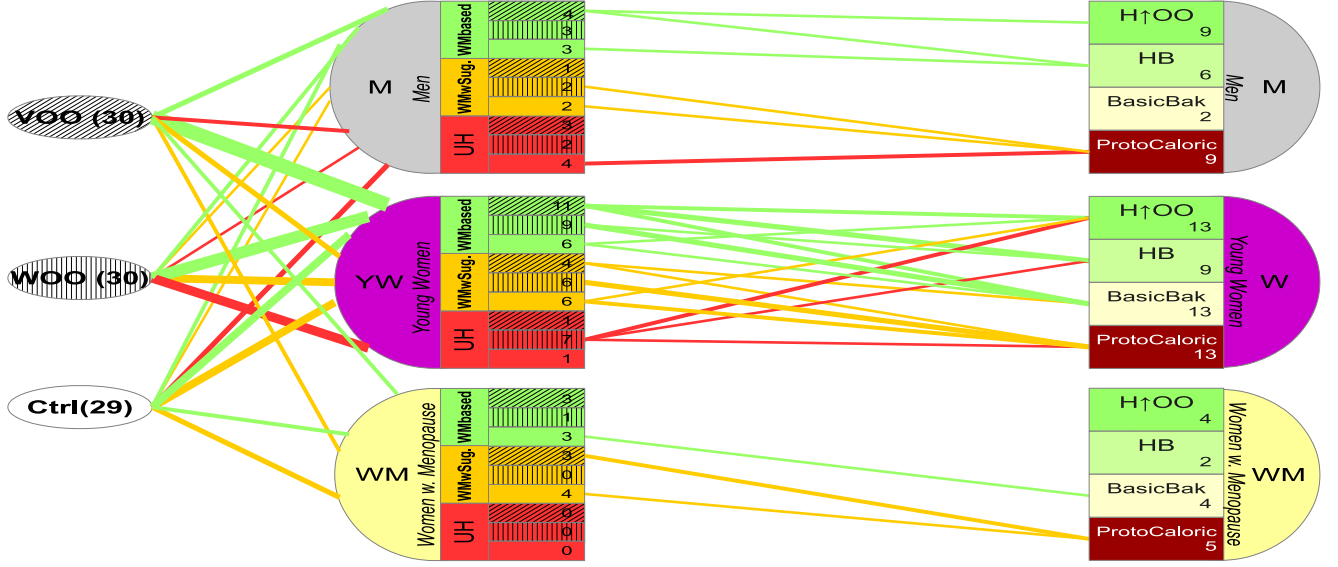
Fig. 2. Analysis of trajectories between the pre ($P_{!o}$) and post ($P_{!f}$) partitions with trajectories with more than one individual.

TABLE III
TABLE OF PROFILE COMPARISONS

Table f_D

	$Class_o$	$Class_f$	Intervention	p14_1 (mainOliveOil)	p14_2 (oliveOil)	p14_3 (vegetables)	p14_4 (fruit)	p14_5 (redMeat)	p14_6 (butter)	p14_7 (gasDrinks)	p14_8 (wine)	p14_9 (legume)	p14_10 (fish)	p14_11 (comBakery)	p14_12 (nuts)	p14_13 (whiteMeat)	p14_14 (Sauce)	Num↑ Healthy	Num↓ Unhealthy	LDL	Triglycerides	F2α Isoprostanes	sP-selectin	sCD40 Ligand	C-Reactive Protein	Tyrosol	ADRB2	ARhGAP15	IFNγ	IL7	POLκ	Num↑ Beneficial	Num↓ Damaging	Size
			VOO	-	-	-	-	-	-	↑	-	-	-	-	↓	-	-	0	5	↓	-	-	-	-	↑	-	↓	↓	↓	↑	↓	0	5	30
			WOO	-	-	-	-	↑	-	↑	-	↑	-	-	-	-	-	1	0	-	-	-	-	-	-	-	-	-	-	-	-	0	0	30
			Ctrl	-	-	↑	-	-	-	-	-	-	↑	-	-	-	-	2	0	-	-	↓	-	↓	-	↓	↑	↑	↑	↑	↑	0	2	29
Trajectory Analysis	M-WMbased	M-HOO	VOO	-	-	-	-	-	-	-	↑	-	-	-	↓	-	-	0	0	-	-	↓	-	-	↑	-	-	-	-	-	-	0	1	2
	M-WMbased	M-HB	VOO	-	-	-	-	-	↑	-	-	-	-	-	-	-	-	0	0	-	-	-	↑	↓	-	-	-	-	-	-	-	0	1	2
	M-WMbased	M-HB	Ctrl	-	-	↑	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	↑	↑	-	-	↓	-	↓	-	↑	0	2	2
	M-WMwSugars	M-ProtoCaloric	WOO	-	-	-	-	↑	-	-	-	-	-	-	-	-	-	0	0	-	-	↓	-	↑	-	↓	-	-	-	-	-	0	1	2
	M-WMwSugars	M-ProtoCaloric	Ctrl	-	-	-	-	-	-	-	↑	-	-	-	-	-	-	1	0	-	-	↓	-	↑	-	-	-	↑	-	-	↑	0	1	2
	M-UH	M-ProtoCaloric	Ctrl	↑	↓	-	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	↑	↓	↑	-	-	-	-	-	-	0	1	3
	YW-WMbased	W-HOO	VOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	↓	↓	↓	↓	↑	↓	↓	↓	↓	↓	1	9	3
	YW-WMbased	W-HOO	Ctrl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	↓	-	↑	↑	↑	↑	↑	-	↑	0	1	2
	YW-WMbased	W-HB	VOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	↓	↓	↓	↓	-	-	↓	-	↑	-	-	0	4	4
	YW-WMbased	W-HB	WOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	↓	↓	↓	↓	-	-	-	-	-	-	0	3	2
	YW-WMbased	W-BasicBak	VOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	↑	↑	↓	↑	↑	-	↓	↑	↑	-	-	-	0	1	3
	YW-WMbased	W-BasicBak	WOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	↓	-	-	↑	-	-	-	-	-	-	-	0	1	4
	YW-WMbased	W-BasicBak	Ctrl	-	-	-	-	-	-	-	-	-	-	↑	-	-	-	0	0	↓	-	↓	-	↓	-	-	-	↓	↑	↓	-	0	5	2
	YW-WMwSugars	W-HOO	Ctrl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	↑	↑	-	-	-	-	-	-	-	0	0	2
	YW-WMwSugars	W-BasicBak	VOO	-	-	-	-	-	-	-	-	↓	-	-	-	-	-	0	0	-	-	-	-	↓	-	-	-	-	-	-	-	0	1	2
	YW-WMwSugars	W-ProtoCaloric	VOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	↓	↑	↑	↑	-	↑	-	-	↑	-	-	1	1	2
	YW-WMwSugars	W-ProtoCaloric	WOO	-	-	-	-	↑	-	-	↑	↓	-	-	-	-	-	1	0	-	↑	-	↑	↑	-	-	↓	↓	↓	-	↓	0	4	4
	YW-WMwSugars	W-ProtoCaloric	Ctrl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	↓	-	-	-	-	↑	↑	↑	↑	↑	0	1	3
	YW-UH	W-HOO	WOO	-	↑	-	-	-	-	↑	-	↑	↓	-	-	-	-	2	1	-	-	-	↓	↓	↓	↓	↑	-	↓	↑	-	0	4	3
	YW-UH	W-HB	WOO	-	-	↑	-	-	-	-	-	-	-	-	-	-	-	1	0	-	-	-	-	↓	-	↓	-	-	-	-	-	0	1	2
	YW-UH	W-ProtoCaloric	WOO	-	-	↑	-	-	↑	-	-	-	-	-	-	-	-	1	0	↑	↑	↓	-	↑	↑	↓	↓	↑	↑	-	↑	0	2	2
	WM-WMbased	WM-BasicBak	Ctrl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	0	0	2
	WM-WMwSugars	WM-ProtoCaloric	VOO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	-	↓	-	-	-	↑	-	-	-	↑	-	0	1	3
	WM-WMwSugars	WM-ProtoCaloric	Ctrl	-	-	-	-	-	-	-	-	-	-	↑	-	-	-	1	0	↑	-	-	-	↑	↑	-	-	-	-	-	-	0	0	2

Top of Table III shows how the global profiles change for each intervention group. There are small changes. Group VOO worsen their habits by consuming more gas drinks and less nuts. Group WOO improves one healthy habit (higher consumption of legumes) while worsens two (higher consumption of red meat and gas drinks); Ctrl group improves two healthy habits (higher consumption of vegetables and fish).

However, this analysis is not using the resulting trajectories. Bottom of Table III shows the profile comparisons of each trajectory. Indeed, bottom of Table III is a refinement of the top of this table. Changes identified in top of Table III for a certain intervention group, in fact, can be attributable to a certain trajectory. For example, the increase of red meat consumption identified for WOO group is due to the individuals of trajectory M-WMwSugars→M-ProtoCaloric and YW-WMwSugars→W-ProtoCaloric and the rest of trajectories of WOO do not really alter significantly this nutritional habit. Thus, making the analysis from this perspective issues other changes that are masked on top of Table III.

In addition, Table III easily allows observing whether the individuals of one trajectory have improved or not through the counters of increased healthy items and decreased unhealthy items. The trajectory from YW-UH → W-H↑OO shows the higher number of positive changes.

Therefore, more focused changes are registered when the differences are analyzed with a finer partition and, adherence to intervention might be evaluated. It is not expected that none of group treated through WOO increase red meat (like

individuals from M-WMwSugars→M-ProtoCaloric do), gas drinks (YW-UH→H-↑OO) or decrease fish (YW-WMwSugars→W-ProtoCaloric) from a Mediterranean diet perspective. Similarly, it is not expected that participants prescribed with VOO increase gas drinks or decrease nuts (like participants from M-WMbased→M-H↑OO do) and decrease legumes (YW-WMwSugars→W-BasicBak).

From Table III, some trajectories assigned to Ctrl intervention have changed their habits. For instance, M-WMbased→M-HB, M-WMwSugars→M-ProtoCaloric, and WM-WMwSugars→M-ProtoCaloric have improved one healthy habit. Nutritional experts outlined that it is common that control group improves their diet because the participants become more aware of its habits by being involved in a clinical trial. In any case, some trajectories do not apparently change their diets during the trial (like YW-WMbased→W-HB, YW-WMbased→W-BasicBak, WM-WMwSugars→WM-ProtoCaloric treated with VOO or WOO). This is, in fact, due to the intrinsic precision of the p-14 questionnaire scale: for example, p14_4 registers whether a person consumes three or more pieces of fruit daily. Thus, all persons with VOO or WOO prescription that moves from no fruit at all before the study to two pieces cannot be identified even if they adopted a better diet than before, and they made greater effort that those moving from two to three who are identified as improving. When more precise information on diet habits be available, the proposed methodology might provide more precise view of the intervention adherence, as well as of diet effect when adhesion. Individual food quantities consumptions would be really welcomed to this purpose, but unfortunately, they were not available by the moment of the submission.

However, it is interesting to contrasts the intake or metabolism of a nutrient with related biomarkers when available. The attribute Tyrosol measures the quantity of polyphenols in urine, a substance that is abundant in virgin olive oil. Then, it is expected that individuals assigned to the VOO intervention increase this level (see Table III on Biomarkers columns). Therefore, analyzing the changes in Tyrosol levels, it is not expected that the groups assigned to VOO have lower values like the ones of YW-WMbased→W-BasicBak. Most of groups assigned to WOO have lower levels and this could be because they have replaced virgin olive oil for washed olive oil which has little polyphenols. Then, the proposal permits identifying irregularities in the adherence to the prescribed diet.

B. Intervention Effects

The same methodology is used to characterize the trajectories using other indicators. Indeed, these changes represent the intervention effects along the study. For this task, only blood/urine biomarkers and gene expression are used because those are the most relevant from a clinical point of view. Table III includes the differences of levels of seven biomarkers and five genes.

There is a complete different scenario of effects depending on each group. For example, individuals of trajectory YW-WMbased → W-H↑OO have a complete different response to the VOO intervention than YW-WMbased (VOO) → W-BasicBak. The first trajectory decreases most of the biomark-

ers levels (F2α Isoprostanes, sP-Selectin, sCD40 Ligand, C-Reactive Protein) and downregulates the five genes (ADBR2, ARhGAP15, Interferonγ, IL7 and POLκ). Whereas, the second trajectory increase LDL, Triglycerides, sP-Selectin, sCD40 Ligand and decreases F2α Isoprostanes and Tyrosol and upregulates ADBR2 and ARhGAP15. Since the semantic of these attributes is available, the first trajectory seems to improve during the intervention, and the second trajectory has worsened.

V. CONCLUSIONS AND FUTURE WORK

This research has been contextualized within the scope of nutritional studies. This work has been motivated by the inherent difficulties of this type of studies when classical approaches were used, as referred in Section I.

The objective of the work is to evaluate both the adherence to the intervention and its effects. The Trajectory Analysis approach has been introduced for this purpose. The characterization of the trajectories allows understanding how the nutritional profiles change in real sample data and along time. This permits specifically analyze the adherence of individuals to assigned interventions, contributing to better identifying the causal relationships between final patterns and individual's characteristics.

To evaluate this analysis, the profiles of the global intervention groups have been compared. In this comparison, more specific changes are observed when the different types of participants are used (trajectories). However, small differences are found because the data does not provide the needed accuracy to assess the adherence to the intervention with precision. For a deeply study of the adherence to diet, a retrospective recovery process is being carried out to obtain a more fine and detailed information about the quantity of each food item. Since the current dietary data comes from a categorical self-reported survey, it is not possible to correctly extract differences.

Besides, the analysis of the effects over each trajectory allows extracting the relationships between intervention effects, type of individuals, and genetic conditions. Thus, further knowledge on the role of genes in metabolizing nutrients can be obtained.

This is an interesting research field since more and more resources are devoted to extract more knowledge for human genome, and there is an increasing activity in the area. The present described strategy will let further personalized analysis of how diet intervention can benefit on each class-group of individuals in the future.

In addition, the analysis of how individuals evolve through the trajectory analysis can be imported to other systems, for example to generate feedback to health professionals. Also, a professional can see the analysis of other aspects such as the biomarkers or gene expression in order to recommend more personalized diets. This information is also valuable for patients who can be aware whether they are following the prescribed diet or not and to adjust the diet themselves. This kind of analysis could be included in an automatic system that generates notifications to the patient or the nutritionist when the adherence is not followed. This automatic system could include food sensors for updating the data in real time.

REFERENCES

[1] National Tech Center, "National Center for Technology Innovation: Experimental Study Design," NCTI at the American Institutes for Research, 2013.

[2] L. Afman and M. Müller, "Nutrigenomics: From molecular nutrition to prevention of disease," *J. Amer. Diet Assoc.*, vol. 106, no. 4, pp. 569–576, 2006.

[3] X. N. Zou *et al.*, "Seasonal variation of food consumption and selected nutrient intake in linxian, a high risk area for esophageal cancer in china," *Int. J. Vitamin Nutrition Res.*, vol. 72, no. 6, pp. 375–382, 2002.

[4] K. Gibert *et al.*, "Knowledge discovery on the response to neurorehabilitation treatment of patients with traumatic brain injury through an AI&Stats and graphical hybrid methodology," *Frontiers Artif. Intell. Appl.*, vol. 184, pp. 170–177, 2008.

[5] Y. M. Chae *et al.*, "Data mining approach to policy analysis in a health insurance domain," *INT J Med. Inform.*, vol. 62, no. 2, pp. 103–111, 2001.

[6] Y.-M. Tai and H.-W. Chiu, "Comorbidity study of ADHD: Applying association rule mining (ARM) to national health insurance database of taiwan," *Int. J. Med. Inform.*, vol. 78, no. 12, pp. e75–e83, 2009.

[7] J. M. Ordovas and V. Mooser, "Nutrigenetics and nutrigenomics," *Casopis Lekaru Ceskych*, vol. 146, no. 2, pp. 837–839, 2007.

[8] V. Konstantinidou *et al.*, "Time course of changes in the expression of insulin sensitivuy-related genes after an acute load of virgin olive oil," *Omics: J. Integrative Biol.*, vol. 13, no. 5, pp. 431–438, 2010.

[9] V. Konstantinidou *et al.*, "In vivo nutrigenomic effects of virgin olive oil polyphenols within the frame of the Mediterranean diet: a randomized controlled trial," *FASEB J.*, vol. 24, no. 7, pp. 2546–2557, 2010.

[10] I. Dahlman *et al.*, "Changes in adipose tissue gene expression with energy-restricted diets in obese women," *Amer. J. Clin. Nutrition*, vol. 81, no. 6, pp. 1275–1285, 2005.

[11] N. Viguerie *et al.*, "Transcriptomics applied to obesity and caloric restriction," *Biochimie* 87, no. 1, pp. 117–123, 2005.

[12] B. Sevilla-Villanueva, "A methodology for pre-post intervention studies: An application for a nutritional case study," Ph.D. thesis, Universitat Politècnica de Catalunya - BarcelonaTech, Barcelona, Spain, 2016.

[13] "Intelligent recommender of personalized diets from an integrated observation of personal characteristics, life-style and health and genetic condition," (Diet4you). Research project funded by Spanish MINECO Ministry (TIN2014-60557-R), 2015–2017.

[14] R. Estruch *et al.*, "Primary prevention of cardiovascular disease with a mediterranean diet," *New England J. Med.*, vol. 368, no. 14, pp. 1279–1290, 2013.

[15] H. Schröder *et al.*, "A short screener valid for assessing Med. diet adherence among older Spanish men and womendel,'" *J. Nutrition*, vol. 141, no. 6, pp. 1140–1145, 2011.

[16] D. A Schoeller *et al.*, "Self-report–based estimates of energy intake offer an inadequate basis for scientific conclusions," Letters to the Editor, *Am. J. Clin.*, vol. 97, no. 6, pp. 1413–1415, Jun. 2013.

[17] B. Sevilla-Villanueva, K. Gibert, and M. Sànchez-Marrè, "Identifying nutritional patterns through integrative multiview clustering," *Frontiers Artif. Intell. Appl.* vol. 277, no. 2, pp. 185–194, 2015.

[18] J. Ward Jr, "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[19] J. A. Gower, "General coefficient if similarity and some of its properties," *Biometrics* vol. 27, pp. 857–874, 1971.

Karina Gibert received the B. Sc M.Sc, and Ph.D. degrees in Computer Science at UPC, in 1990, 1991, 1995, respectively.

She is an Associate Professor at the UPC. Her main research interests include hybrid AI and Statistics data mining, pre and post-processing, intelligent decision support, data science, with focus to multidisciplinary and applied research (vast experience in health, environmental systems, e-commerce or tourism), as well as technology transfer to productive sector. Dr. Gibert is Pioneer member (1986) of the KEMLG research Group, member of the IASC Committee on Computational Statistics and Data Mining for Knowledge Discovery, Board member of the International Environmental Modelling and Software Society (iEMSs). She is Pioneer member of several scientific associations related to Data Mining, Statistics and AI. She is a Fellow of iEMSs. She was researcher of project K4Care (EU-best e-Health project, 2010), PI of Diet4You Spanish project (2015-2017) and GENESIS project (winner of 1rst HackingBullipedia contest). Intensive Scientific production with more than 300 papers (H-index: 20). Additional details can be seen at www.eio.upc.edu/homepages/karina.

Miquel Sànchez-Marrè received the B.Sc. and M.Sc. degrees in Computer Science both from the Barcelona School of Informatics, in 1988 and 1991, respectively. He received a Ph.D. in Computer Science in 1996 from the Universitat Politècnica de Catalunya- BarcelonaTech.

He is Associate Professor in the Computer Science Department (CS) of the UPC since 1990 (tenure 1996). His main research topics are case-based reasoning, machine learning, data mining, recommender systems, data science, knowledge discovery, knowledge engineering, intelligent decision support systems, and integrated AI architectures. He is a member of the Editorial Board of International Journal of Applied Intelligence and an Associate Editor of the *Environmental Modelling & Software journal*. He is a Fellow of the iEMSs since 2005 (H-index:22).

Montserrat Fitó received the Ph.D. degree in Medicine.

She is a Specialist in clinical analyses. She is the Head of the Cardiovascular Risk and Nutrition Research Group (CARIN) in the Institut Hospital del Mar d'Investigacions Mèdiques (IMIM) and also leader of this group in the Centro de Investigación Biomédica en Red (CIBER) (Physiopathology of Obesity and Nutrition Area-CIBEROBN), Madrid (Spain). She is the Technical Director of the REGICOR Group Laboratory (from CARIN) and Genetic and Cardiovascular Epidemiology Research Group (EGEC), in the IMIM. Intensive research activity (H-index: 35).

María-Isabel Covas received the degrees in pharmacy and biology, and the Ph.D. degree in biochemistry.

She was a Scientific Adviser of NUPROAS Handesbolag (2013-act) and Head of CARIN group (IMIM, 1999–2013). She was an Associated Professor at Barcelona University (1974–1983), the Deputy Head of the Clinical Laboratories of the Hospital del Mar and Hospital de la Esperanza (1983–1995), a Professor at Medicine School, Barcelona Autonomic University (1988–2004), a Professor in Clinical Nutrition of Barcelona University (2002–2012). Her main research interests include nutrition and protective factors for cardiovascular disease, with a special focus on the Mediterranean diet, olive oil, wine, and their polyphenols. Intensive Scientific production with more than 200 papers in international journals. She was a member of the Scientific Board of Spanish and European Institutions related with Nutrition (2004–2012). Awards of Excellence in Nutrition Research (AEMO (2003); Carles Mart Heinneberg from the Spanish Danone Institute (2009); and Ramón Turró from the Catalan Center of Nutrition (2015).

Beatriz Sevilla-Villanueva received the B. Eng. in informatics (2007) and Ms.C degree and Ph.D in Artificial Intelligence, in 2009 and 2016, respectively, from the Universitat Politècnica de Catalunya-BarcelonaTech (UPC), Barcelona, Spain.

She is a researcher at the UPC. Involved in several European Projects (2006-2013) and currently working in a Spanish Project Diet4You. Her research interests include AI, data mining, nutrition, nutrigenetics and nutrigenomics, clustering, cluster interpretation, case-based reasoning and multidisciplinary and applied research.

Dr. Sevilla-Villanueva is a member of the Knowledge Engineering and Machine Learning Group (KEMLG) (excellence certificate from Catalan government) since 2006 and a member of the ACIA (Catalan Association of Artificial Intelligence). She was a researcher of the project GENESIS, which won the Hacking-Bullipedia contest (Telefonica and elBulliFoundation from the famous chef Ferran Adrià, November 2013)