Categorical state sequence analysis and regression tree to identify determinants of care trajectory in chronic disease: example of end-stage renal disease

Nolwenn Le Meur[*1], Cécile Vigneau[2,3], Mathilde Lefort[1], Saïd Lebbah[4], Jean-Philippe Jais[4], Eric Daugas[5], Sahar Bayat[1]

(1) Univ Rennes, EHESP, REPERES (Recherche en Pharmaco-épidémiologie et Recours aux Soins) - EA 7449, F-35000 Rennes, France

(2) CHU Pontchaillou, Service de Néphrologie, Rennes, France

(3) IRSET, INSERM UMR 1085, Rennes, France.

(4) CHU Necker Enfants Malades, Biostatistics unit, INSERM UMR S 872, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

(5) Hôpital Bichat, Service de Néphrologie, DHU FIRE, INSERM U1149, Université Paris Diderot, Paris, France

* Corresponding author: Nolwenn Le Meur

Affiliation: *Univ Rennes, EHESP Rennes, Sorbonne Paris Cité, Département METIS, REPERES Pharmacoepidemiology and Health Services Research - EA 7449*

Postal address: EHESP, Avenue du Professeur Leon Bernard, 35043 Rennes, FR

E-mail: nolwenn.lemeur@ehesp.fr

Phone: +33 2 99 02 25 14

ABSTRACT

Background

Patients with chronic diseases, like patients with end-stage renal disease (ESRD), have long history of care driven by multiple determinants (medical, social, economic...). Although in most epidemiological studies, analyses of health care determinants are

computed on single health care events using classical multivariate statistical regression methods. Only few studies have integrated the concept of treatment trajectories as a whole and studied their determinants.

Methods

All 18 to 80-year-old incident ESRD patients who started dialysis in Ile-de-France or Bretagne between 2006 and 2009 and could be followed for a period of 48 months after initiation of a renal replacement therapy were included (n= 5 568). Their care trajectories were defined as categorical state sequences. Associations between patients' characteristics and care trajectories were assessed using a regression tree model together with a discrepancy analysis.

Results

On average, each patient experienced 1.56 different renal replacement therapies (min=1; max=5) during the 48 months of follow-up. About 55% of patients never changed treatment and only 1% tried three or more renal replacement therapy modalities. Twelve homogeneous care trajectory groups were identified. Covariates explained 12% of the discrepancy between groups, particularly age, regions and initiation of hemodialysis with a catheter.

Conclusions

Regression tree analysis of categorical state sequence highlighted geographical disparities in the care trajectory of French patients with ESRD that cannot be observed when focusing on a single outcome, such as survival. This method is an original tool to visualize and characterize care trajectories, notably in the context of chronic condition like ESRD.

Keywords: categorical state sequence; regression tree; discrepancy; care trajectory; end-stage renal disease.

INTRODUCTION

Patients with chronic diseases have long history of care driven by multiple determinants (medical, social, economic...). For instance, end-stage renal disease (ESRD) patients need renal replacement therapy (RRT). The treatment modalities (dialysis or renal transplantation) are chosen in function of each patient's characteristics and the local healthcare offer[1]. During his/her care trajectory, a ESRD patient may undergo different RRT modalities to which are added interacting events (e.g., registration to the transplantation wait-list, temporal inactive status) that could affect morbidity, mortality and quality of life[2]. Better understanding the care trajectories of ESRD patients and their determinants is therefore an epidemiological issue.

Currently most epidemiological studies of ESRD patients have analyzed the effect of one RRT modality (compared to a reference) on a given outcome (e.g., access to transplantation, graft survival or death) using classical multivariate statistical regression methods. Only few studies have integrated the concept of treatment trajectories with the aim of predicting the future healthcare demands based on a few determinants as age or diabetes status[3,4]. Moreover these studies accounted for determinants independently[5–8] and did not integrate the potential sequence of treatments along with the possible interactions between determinants that compose a patient care trajectory.

To decipher ESRD patient care trajectory, we propose to demonstrate the use of regression tree analysis applied to categorical state sequence objects (CSS) as published by Studer *et al.*[9] First developed in social sciences to characterize individual life trajectories (e.g. occupational history, professional career or cohabitation life courses), categorical sequences analysis allows transversal aggregated views of longitudinal data such as medical care trajectories. In health services research, we demonstrated that clustering analysis of care trajectory defined as CSS allowed identifying homogenous

groups of care trajectories among pregnant women that were partly explained by their socio-economical environment[10]. Because cluster-based approaches consist on reducing the set of sequences to a limited number of standard trajectories, one loses information about the diversity within each cluster[9]. Furthermore, knowledge of the cluster membership alone does not inform about the distances and differences between clusters. Therefore as proposed by Studer *et al.*[11], we suggested to use discrepancy analysis of CSS associated with a regression tree method to quantifies the amount of variation between medical care trajectories (sequences) that can be explained by covariates or their interactions. Discrepancy among care trajectories can be derived from their pairwise dissimilarities, which permits then to identify factors that most reduce this discrepancy. In a public health context, this approach could help showing how care trajectories are related to covariates, such as patient characteristics, healthcare offer or geographical location. The latest is of importance in France as two studies already demonstrated that after taking into account medical and non-medical factors, placement on the kidney transplant waiting list is significantly associated with the region of residence, suggesting wait-listing practice differences[12,13].

In this paper, we (i) propose the use regression tree model on CSS as an original tool to characterize and visualize care trajectories, and (II) demonstrate its application on ESRD patient care trajectories by identifying groups of patients of homogeneous care trajectories explained by a mixture of covariates.

MATERIALS AND METHODS

Data

Data were extracted from the French Renal Epidemiology and Information Network (REIN) registry[14]. All 18 to 80-year-old incident patients with ESRD who started dialysis in Ile-de-France or Bretagne between 01/01/2006 and 12/31/2009 were included. ESRD patients

older than 80 years of age were excluded because of their extremely low probability of access to the waiting list and transplantation. For comparison purposes, selected patients had to be followed for a duration of 48 months after the first RRT. Patients who underwent RRT for less than 48 months, received preemptive graft, moved to other regions, or were lost to follow-up were excluded as their treatment data were not available. The different RRT modalities were: in-center hemodialysis [CENTER], hemodialysis in a medical satellite unit [MSU], autonomous hemodialysis (self-care unit and home) [AUTO], peritoneal dialysis [PD], and graft [GRAFT].

Potential determinants of care trajectories included (i) three demographic covariates: age group (18-39, 40-59, 60-69 and 70-80 years old), sex, and region of residence (Bretagne or Ile-de-France) (ii) clinical features at first dialysis: the initial renal disease, categorized in six groups (glomerulonephritis, pyelonephritis, diabetic nephropathy, hypertensive and vascular nephropathy, autosomal dominant polycystic kidney disease, and other causes or unknown), emergency first dialysis and first dialysis session with a catheter (1/0), serum hemoglobin levels in 3 classes (<10 g/dl, [10-12] g/dl and >12 g/dl), presence of cardiovascular diseases (0,1, 2 or ≥3), diabetes (1/0), chronic respiratory disease (1/0), hepatic disease (1/0), active malignancy (1/0), physical disabilities (ambulation impairment, paraplegia or hemiplegia, member amputation, blindness; coded as ≥1 or 0), and HIV infection (1/0). Quantitative covariates were grouped into classes, as for age and hemoglobin levels. Binary variables were coded 1 and 0 for presence or absence of the characteristic for the patient, respectively.

Care trajectory definition

A care trajectory was composed by ordered suite of RRT modalities that defined a CSS $X = (x_1, x_2, \ldots, x_n)$, where x_i is a RRT modality (Figure 1). Two extra states were used to indicate the registration in the transplantation waiting list [WAITLISTED] and the death

[DIED]. Death was forced as a state to better disentangle which patients' characteristics drive care trajectory whatever the health issue (i.e. survival or death) within the 48 months of follow-up. All calendar dates were converted into time intervals in month units starting at the first RRT and ranging from 1 to 48 months. State sequences of patients from Bretagne were given a weight of 4 to account for the 1:4 ratio difference in population size between Bretagne and Ile-de-France.

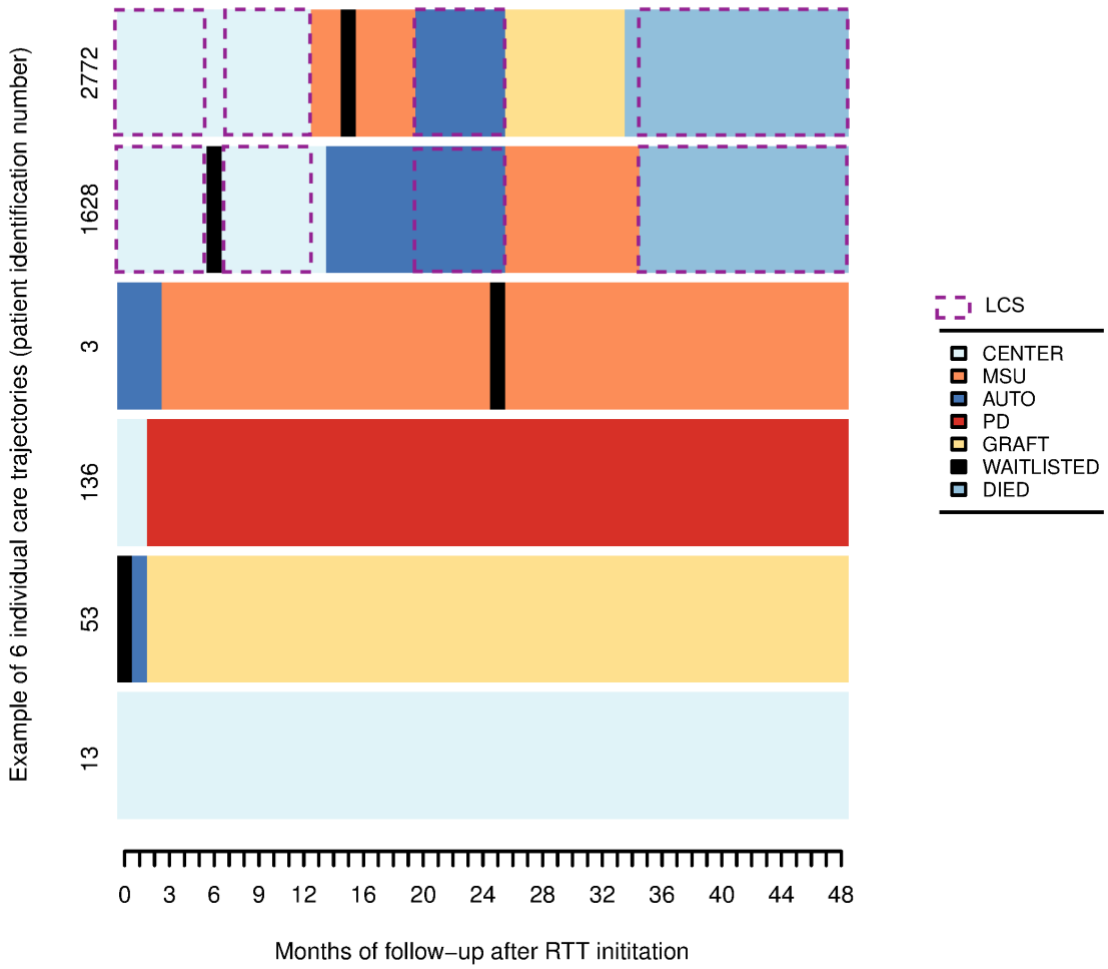


Fig. 1 Trajectory of care as an ordered sequence of states. For each ESRD patient, the succession of renal replacement therapies (RTT) was translated into an ordered sequence of states for a 48-month period after RRT initiation. The duration of each state was the duration of each treatment phase. The death event was interpreted as an irreversible state. The event of being registered to the transplantation waiting list was defined as a punctual state with 1-month duration. Overall seven states were possible: in-center hemodialysis [CENTER] in aqua blue, hemodialysis

in medical unit [MSU] in orange, autonomous hemodialysis [AUTO] in dark blue, peritoneal dialysis [PD] in red, death [DIED] in grey blue, transplantation [GRAFT] in yellow, registration point to the transplantation waiting list [WAITLISTED] in black. For example, ESRD patient #2772 initiated his treatment by in-center hemodialysis for 13 months after which he started hemodialysis in medical unit for 7 months and in the meantime he was registered to the transplantation waiting list (16 months after initiation of his RRT). After hemodialysis in medical unit, he received a graft at the 26th month but died 8 months later (34 months after RTT initiation). For patients #2772 and #1628, the dashed purple rectangles highlight the common sub-sequences taken into account by the LCS algorithm to compute their similarity (and the dissimilarity).

Discrepancy measure and Regression tree for CSS

The regression tree was built following the ANOVA principles. CSS were recursively partitioned in two subgroups based on state sequence–covariate associations that minimize within-group discrepancies and maximize between-group variations.

Discrepancies (noted S^2) were determined using pairwise dissimilarities between patient's state sequences (Figure 1). Pairwise dissimilarities were computed using the longest common subsequence (LCS) distance algorithm as defined and available in the R library *TraMineR* (version 1.8.13)[15]. LCS is a metric based on counts of common attributes occurring in the same order between state sequences[16]. Given two sequences $X = (x_{1,}x_2, \dots, x_n)$ and $Y = (y_{1,}y_2, \dots, y_n)$, Z is a common subsequence of X and Y if it is a subsequence of both X and Y. Z is a longest common subsequence (LCS) of X and Y if $|Z| \geq |Z'|$, for all other common subsequences Z' of X and Y (Figure 1). Among the sequence dissimilarity measures available in *TraMineR,* LCS metric was chosen as is not too sensitive to timing but it presents a dependence on the sequencing and the duration spent in the distinct states, which matter most in long term RRT. Optimal matching (OM) measures[18], frequently used in social sciences, was less justified in our context since short time deletion or substitution transformations had less meaning[17]. The care trajectory of one RRT patient could hardly be considered to be the result of a transformation of a care

trajectory of another patient.

The sum of weighted distances (SS) between care trajectories was then computed following Equation (1):

$$SS = \frac{1}{W}\sum_{i=1}^{n}\sum_{j=i+1}^{n} w_i\, w_j d_{ij} \qquad (1)$$

where *n* is the number of cases, w_i and w_i are the weights associated to case *i* and *j* respectively, W is the total sum of weights, and d_{ij} is the LCS distance between *i* and *j*. The variance or, in more general term, the discrepancy S^2 (for non-Euclidian distances as LCS) was then $S^2 = \frac{1}{W}SS$, which is equal to half of the weighted average of the pairwise dissimilarities (Equation 2).

$$S^2 = \frac{1}{2W^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n} w_i\, w_j d_{ij} \qquad (2)$$

The regression tree growing process started with the entire study sample (root). Recursively, the tree-generating algorithm tested all values of the covariates to split the original set in two subgroups (left and right daughters) that minimize within-group discrepancies and maximize between-group variations (as standard performed for regression tree CART[19,20]). More precisely, at each node, each covariate and their possible binary splits of category levels were tested and the split with the highest univariate pseudo-R^2 was selected to create two child nodes. The part of the discrepancies explained by a covariate (pseudo-R^2) was quantified using the Huygens theorem of the ANOVA approach. The theorem states that the total sum of squares (SS_T) is the between (groups) sum of squares (SS_B) plus the residual within (group) sum of squares (SS_W). SS_T and SS_W are computed from Equation (1). SS_W is the sum of the within sums of squares of each subgroup. SS_B is then the difference between SS_T and SS_W. It follows that $R^2 = \frac{SS_B}{SS_T}$ is the proportion of total discrepancy explained by a covariate. Alternatively, $F = \frac{SS_B/(m-1)}{SS_W/(W-m)}$ is the

explained discrepancy to the residual discrepancy where *m* is the number of groups and *W* the total sum of weights. The assessment of the statistical significance of the split (or strength of the association between trajectories and covariate) was performed using permutation tests on the *F* statistics. Within each subgroup, the *F* statistics was computed for the observed values of the covariates of each sequence, namely F_{obs}. At each permutation, the value of the covariate was assigned randomly to each sequence and a F_{perm} value was computed. Five thousand permutations and a *p*-value threshold of 5% were used to assess the split significance. The p-value was the proportion of F_{perm} values that were higher than F_{obs}. The regression tree growing process stopped when empirical stopping criteria were reached. We gave the arbitrary constraint that the covariates of the last partitioning step should explained at least 1% of the discrepancy between subgroups which in our case corresponded to a maximum depth of 5 leaves in the regression tree. Only covariates that contributed significantly to partitioning the observed trajectories were selected (R^2 >= 1% and significant *F* permutation test).

The global quality of the tree was assess using the *pseudo-F* and Levene's tests that summarize the statistical significance of the obtained segmentation compared with random splits for the between-groups and within-group discrepancies, respectively. The global *pseudo-*R^2 was calculated as a measure of the part of the total discrepancy that was explained by the regression tree.

All computation were performed using R (version 3.3.3 - R Foundation, Vienna, Austria)[21], and the *TraMineR* library (version 1.8.13)[15,22].

RESULTS

Treatment sequences during the first 48 months of RRT

We retrieved data for 6166 patients (5045 in Ile-de-France and 1121 in Bretagne). Among

them, 905 were excluded because they moved to other regions or were lost to follow-up (n=357), could not be followed for 48 months (n=273), or received preemptive graft (n=310). Therefore, 5568 patients were included in the analysis (Table 1).

Table 1. Characteristics of the adult ESRD patients registered in REIN between January 1, 2006 and December 3, 2009 and followed for 48 months before the study end point (December 31, 2013). (Exclusion criteria: preemptive graft, lost to follow-up, and moved to other regions)

	Bretagne N= 1062	**Ile-de-France N=4506**	**All patients N = 5568**
Age (mean ± SD) in years	64.3 ±14	59.9± 15	60 ± 15
(min – max)	(19 – 80)	(18 – 80)	(18 – 80)
Sex ratio (M/F)	1.57	1.80	1,75
Registered in the transplant waiting list (%)	357 (34)	2203 (49)	2560 (46)
Transplantation within 48 months (% of registered)	320 (90)	1182 (54)	1502 (59)
Died within the 48 months of follow-up (%)	396 (37)	1438 (32)	1834 (33)

Although care trajectories seemed heterogeneous with a diversity of switches between RTT, 72% of ESRD patients had their first RRT in-center hemodialysis (Figure 2 and Figure S1). Over the 48-month follow-up, the proportion of patients undergoing peritoneal dialysis and in-center hemodialysis decreased (essentially in favor of transplantation or because of death), while the percentage of patients undergoing hemodialysis in a medical satellite unit increased from 1% to 6%. The observation was similar in both regions (Figure S2).

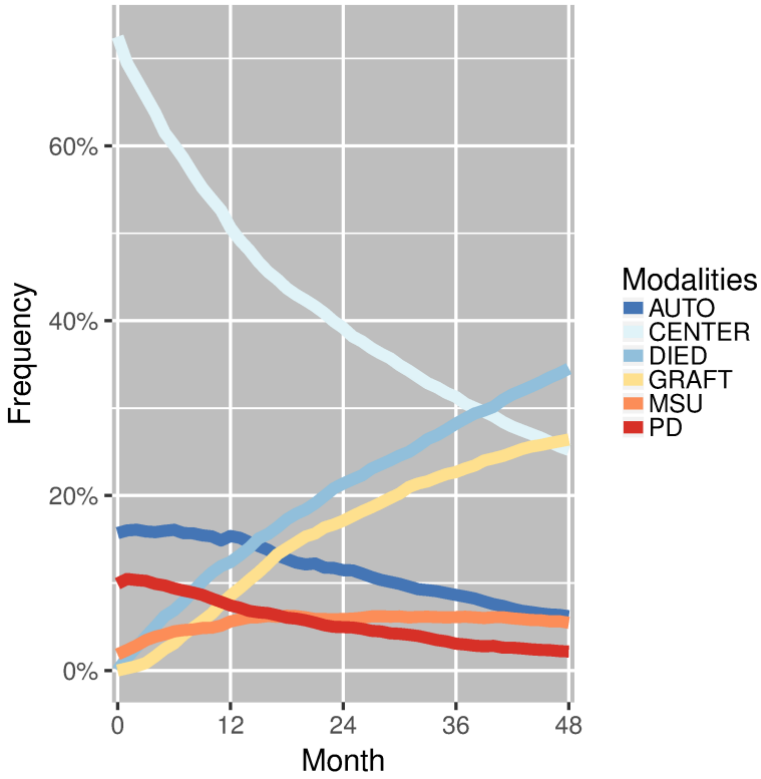


Fig. 2 Percentage of ESRD patients for each RRT modality per month during the 48-month follow-up period. At initiation around 72% of the patients were in hemodialysis in center and at the end of the study period around 25 % of the patients were take care of in center, 35% died, and 25% received a graft. AUTO: autonomous hemodialysis; CENTER: in-center hemodialysis; MSU: hemodialysis in medical unit; PD: peritoneal dialysis.

On average, a patient underwent 1.56 different treatments (min=1; max=5) during the 48 months of follow-up. About 55% of patients never changed RRT modality, 44% changed once or twice, and only 1% changed three or more times. Patients who tried two or more RRT modalities generally switched between the same treatments, for instance, in-center hemodialysis, then autonomous hemodialysis followed again by in-center hemodialysis and then autonomous hemodialysis, and so on (Figure S1). A patient underwent 1.64 switches on average, with a maximum of 8. When switches occurred, the most frequent transitions were from in-center hemodialysis to autonomous dialysis, from hemodialysis in a medical satellite unit to in-center hemodialysis, and from autonomous dialysis or peritoneal dialysis to graft. Over the 48 month-period after RRT initiation, patients spent on

average 25 months on in-center hemodialysis, 22 months on autonomous dialysis, 20.7 months on hemodialysis in medical satellite units, or 24.5 months on peritoneal dialysis before switching to another modality (including graft), or death. Patients waited on average 8.8 months (median: 6 months) before being placed on the transplantation waiting list. Among waitlisted patients (n=2560), 20% were registered at RRT initiation, 25% within 2 months, and 75% within the first year of RRT. Overall, 58% of waitlisted patients underwent kidney transplantation, after a mean time on the waiting list of 25 months. At the end of the 48-month follow-up, 1834 patients (33%) were dead and 50% of them within the first 18 months after RRT initiation.

Determinants of homogeneous groups of care trajectory

Twelve subgroups of homogeneous RRT trajectories were identified (Figure 3). The trajectories varied significantly more between groups than within groups (based on the results of the Levene's and pseudo F tests). The tree explained about 12% of the total discrepancy observed. A path from the root to the terminal leaf corresponded to the effects of interactions between the respective variables on discriminating patients' care trajectories. For instance, on the left side of the tree from top to bottom, patients were younger than 60 years of age AND lived in Bretagne AND underwent a first dialysis session without a catheter AND had glomerulonephritis or autosomal dominant polycystic kidney disease as primary renal disease.

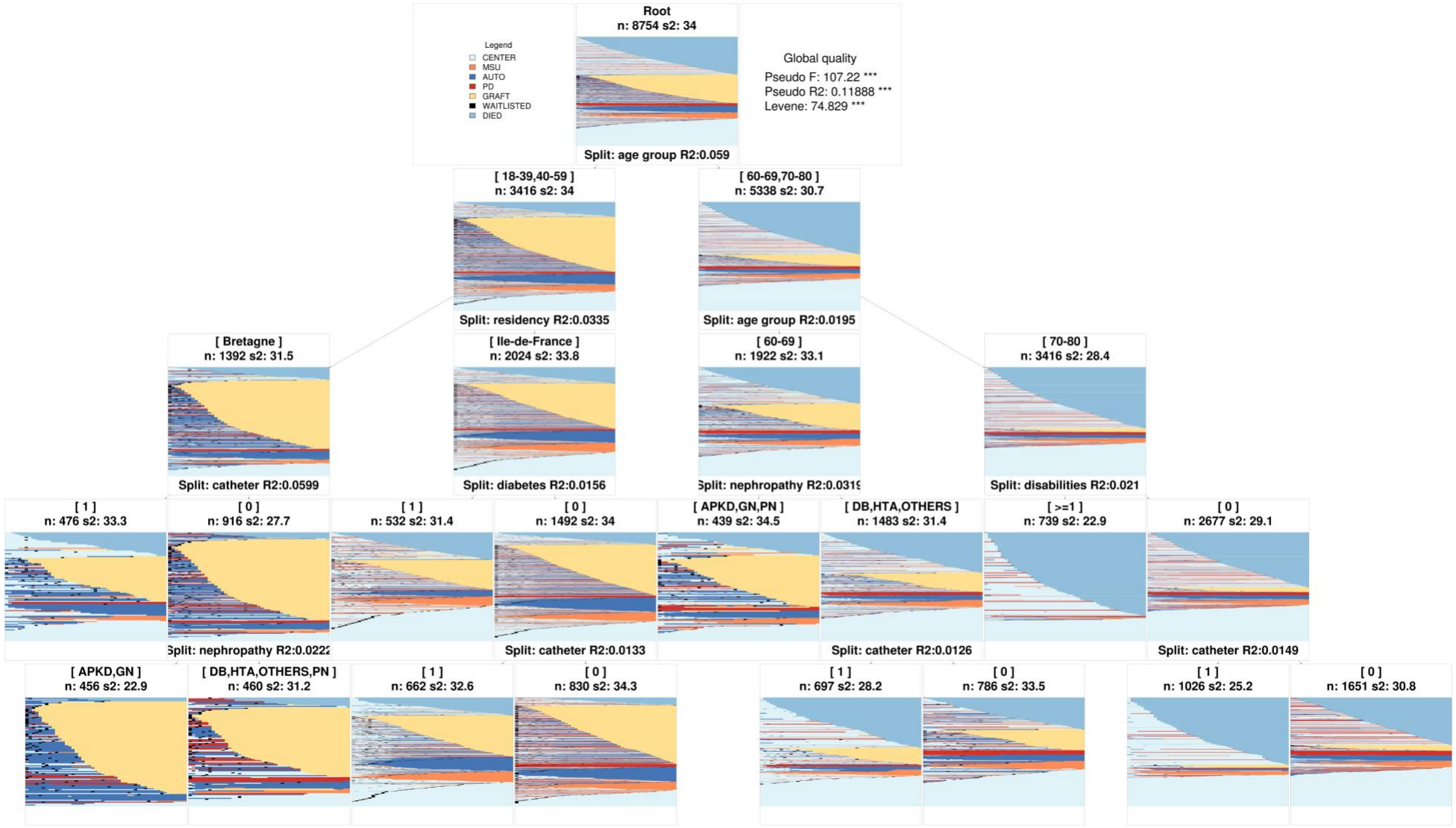

Legend
CENTER
MSU
AUTO
PD
GRAFT
WAITLISTED
DIED
Root
n: 8754 s2: 34
Split: age group R2:0.059
Global quality
Pseudo F: 107.22 ***
Pseudo R2: 0.11888 ***
Levene: 74.829 ***
[18-39,40-59]
n: 3416 s2: 34
Split: residency R2:0.0335
[60-69,70-80]
n: 5338 s2: 30.7
Split: age group R2:0.0195
[Bretagne]
n: 1392 s2: 31.5
Split: catheter R2:0.0599
[Ile-de-France]
n: 2024 s2: 33.8
Split: diabetes R2:0.0156
[60-69]
n: 1922 s2: 33.1
Split: nephropathy R2:0.0319
[70-80]
n: 3416 s2: 28.4
Split: disabilities R2:0.021
[1]
n: 476 s2: 33.3
[0]
n: 916 s2: 27.7
Split: nephropathy R2:0.022
[1]
n: 532 s2: 31.4
[0]
n: 1492 s2: 34
Split: catheter R2:0.0133
[APKD,GN,PN]
n: 439 s2: 34.5
[DB,HTA,OTHERS]
n: 1483 s2: 31.4
Split: catheter R2:0.0126
[>=1]
n: 739 s2: 22.9
[0]
n: 2677 s2: 29.1
Split: catheter R2:0.0149
[APKD,GN]
n: 456 s2: 22.9
[DB,HTA,OTHERS,PN]
n: 460 s2: 31.2
[1]
n: 662 s2: 32.6
[0]
n: 830 s2: 34.3
[1]
n: 697 s2: 28.2
[0]
n: 786 s2: 33.5
[1]
n: 1026 s2: 25.2
[0]
n: 1651 s2: 30.8

Fig. 3 Determinants of homogeneous groups of RRT sequences among ESRD patients. In each panel, each row represents one patient care trajectory over the 48-month follow-up, sorted according to the final state (treatment modality). The RRT modalities can be in-center hemodialysis [CENTER] in aqua blue, hemodialysis in medical unit [MSU] in orange, autonomous hemodialysis [AUTO] in dark blue, peritoneal dialysis [PD] in red, death [DIED] in grey blue, transplantation [GRAFT] in yellow, registration point to the transplantation waiting list [WAITLISTED] in black. Pseudo R2 or R2 shows the proportion of explained variance by the covariate. S2 is the variance of residuals. The pseudo F provides the statistical significance of the segmentation. The Levene's test gives the significance of variance equality within groups/panels. APKD: autosomal dominant polycystic kidney disease; catheter: first dialysis with catheter; DN: diabetic nephropathy; GN: glomerulonephritis; HTA: hypertensive and vascular nephropathy; OTHERS: other causes or unknown; PN: pyelonephritis.

The first determinant was age that contributed to explain almost 6% of the differences between groups. The covariate “first dialysis session with a catheter” significantly contributed to more than one split, but at different levels. Among patients younger than 60 years of age in Bretagne, it accounted for 6% of discrepancy and seemed to separate patients who mainly underwent in-center hemodialysis (right side) from those on autonomous dialysis. Similarly, the primary renal disease contributed to two splits: one among the 60 to 69-year-old patients and one among those younger than 60 years old. The grouping of the different modalities varied between splits. Among patients above 70 years old, physical disabilities were the primary determinant that grouped patients who underwent in-center hemodialysis, were rarely waitlisted, did not have access to kidney transplant and displayed high mortality. Among patients under 60 years old, the region of residency was the first determinant and accounted for 3% of all discrepancies. The Ile-de-France sub-group was characterized by a majority of patients starting RRT with in-center hemodialysis (75%) (Figure 3 and Figure S3). Conversely, in the Bretagne subgroup, equivalent proportions of patients started with in-center hemodialysis and autonomous dialysis (45% and 42%, respectively). In addition, by the end of the 48-month follow-up period, a higher proportion of patients had access to kidney transplantation in Bretagne (63%) than in Ile-de-France (42%) (Figure 3 and Figure S4). In Bretagne, the last two splits were driven by the “first dialysis session with a catheter” and the “primary renal disease” variables. Specifically, 85% of patients under 60 years old in Bretagne with autosomal dominant polycystic kidney disease or glomerulonephritis and who did not receive their first dialysis session with a catheter were transplanted within the study period (versus 63% for the other nephropathies). In Ile-de-France, diabetes contributed to the care trajectory discrepancies more significantly than initiation of dialysis with a catheter. In both regions, young patients (and without diabetes in Ile-de-France) who received a first dialysis session with a catheter were less likely to underwent kidney transplantation than

those who did not.

DISCUSSION

While analysis of health care determinants is often computed on single health care events, we believe that monitoring care trajectories and identifying their determinants offer opportunities to highlight point of leverage for improvement of long term care[10]. The formalization of care trajectories as CSS leads to an interesting and synthetic longitudinal display that allows pattern identification through cluster analysis, state transition monitoring, or trajectory analysis[3,10,23]. Regression tree method are an innovative alternative strategy to descriptive and cluster-based analyses of CSS. As cluster-based approaches reduce the set of sequences to a limited number of standard trajectories, information about the diversity within each cluster is lost[9]. Furthermore, knowledge of the cluster membership alone does not give information about the distances and differences between clusters. Regression tree methods overcome these limits and allow the direct analysis of the sequence-covariate links and the measurement of the relationship strength. These methods take into account the inter-individual variability of the trajectories, while studying the relationships between the trajectories and their contexts.

This analysis of the REIN registry on ESRD patients who lived in two different regions shows care trajectory variations and the contribution of demographical, medical, and regional covariates to these differences. Though life courses of ESRD patients are specific and health care providers proposed personal care, ESRD patients had rather stable sequence of treatments and could be classified in homogenous sub-groups. Whatever the region, ESRD patients had simple care trajectories with few treatment changes. Unsurprisingly, our study shows that the main determinant of care trajectories is age. Whatever the region, patients older than 60 years old were less frequently on the

transplant waiting list and underwent kidney transplant more rarely than younger patients[24]. Initiation of hemodialysis with a catheter is another predictable determinant of care trajectory differences. This medical procedure is associated with early mortality[25,26]. This procedure does not increase the mortality risk per se (except in the rare case of infections), but it is a marker of a more complex RRT initiation at any age. It is often proposed to less compliant patients, critically ill patients, or patients not referred by a nephrologist and who need RRT in emergency. In term of care trajectories, few of these patients were waitlisted and they predominantly used in-center hemodialysis, whatever the age. Most surprisingly, region of residence is the first determinant of care trajectory among patients younger than 60 years old, before all clinical characteristics. This could be explained by regional variations in the RRT modality offer and choice. Although several RRT management changes have recently been implemented in France with the aim of developing alternatives to in-center hemodialysis to improve the quality of life of young and active patients[24], in-center hemodialysis remains predominant, especially at initiation (Figure S2 and S3). These regional variations could also be partially explained by the higher proportion of diabetic patients in Ile-de-France. These patients have complex care trajectories. They are less likely to be registered on the transplant waiting list[12] and to have access to kidney transplantation[13].

One limitation of our result might be the small proportion of the total discrepancy explained by the computed tree. Overall, 12% of the variation observed between care trajectories was explained by the regression tree. However, as far as human behavior is involved, this level is not surprising. In the original work by Studer *et al.* [9], socio-economic and demographic variables contributed to explain 18% of the school-to-work transition in Northern Ireland. Similarly, another group showed that people activity, age, and household size explain 19% of people's daily activity-travel patterns[27]. Moreno-Black *et al.*[28] (2016)

showed that 3% of the variation of body mass index trajectories in elementary school children is explained by low socio-economic status and ethnicity. In our study, we identified homogeneous groups of care trajectories, but the within-group variation remained important. One explanation might be that although patients within a homogenous group of care trajectory used the same RRT modality with few switches, they switched between modalities at different points in time, thus increasing the variability between individuals. Other factors not collected by REIN registry could also have influenced the switches between RRT modalities. For instance, RRT choice could be guided by the type of transportation available to the patients or the travel time between home/work place and dialysis center. Nevertheless, the regression tree analysis of CSS has the advantage to estimate the association and interaction of multiple factors along with the RRT modality sequence.

In conclusion, regression tree analysis of CSS is an original tool to visualize and characterize care trajectories in the context of chronic disease. In health services research and personalized medicine, modelling care trajectory as a whole could contribute to better assess the impact of new treatment modality or the restructuration of health care facilities on patients care trajectory[3]. In addition, this mixed method between data mining and regression could be a promising tool for the exploration of medico-administrative data or claim data that by construction collect valuable information on healthcare consumption (e.g. drug consumption, medical device usage or visits to health care professional and health care facilities) and allow care trajectory analysis[29,30].

FUNDING: This study was supported by the Agence Nationale de la Recherche in the framework of the “Investment for the Future” program [grant number ANR-11-IDEX-0005-02].

CONFLICT OF INTEREST: The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENT: We would like to acknowledge Cécile Couchoud for organizing the extraction of the data from the REIN registry.

REFERENCES

1. Robinson BM, Akizawa T, Jager KJ, et al. Factors affecting outcomes in patients reaching end-stage kidney disease worldwide: differences in access to renal replacement therapy, modality use, and haemodialysis practices. *The Lancet* 2016; 388: 294–306.

2. Joshi VD. Quality of life in end stage renal disease patients. *World J Nephrol* 2014; 3: 308.

3. Couchoud C, Dantony E, Elsensohn M-H, et al. Modelling treatment trajectories to optimize the organization of renal replacement therapy and public health decision-making. *Nephrol Dial Transplant* 2013; 28: 2372–2382.

4. Roderick P, Davies R, Jones C, et al. Simulation model of renal replacement therapy: predicting future demand in England. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc* 2004; 19: 692–701.

5. Fritsche L, Vanrenterghem Y, Nordal KP, et al. Practice variations in the evaluation of adult candidates for cadaveric kidney transplantation: a survey of the European Transplant Centers. *Transplantation* 2000; 70: 1492–1497.

6. Stel VS, van Dijk PCW, van Manen JG, et al. Prevalence of co-morbidity in different European RRT populations and its effect on access to renal transplantation. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc* 2005; 20: 2803–2811.

7. Ravanan R, Udayaraj U, Ansell D, et al. Variation between centres in access to renal transplantation in UK: longitudinal cohort study. *BMJ* 2010; 341: c3451.

8. Oniscu GC, Schalkwijk AAH, Johnson RJ, et al. Equity of access to renal transplant waiting list and renal transplantation in Scotland: cohort study. *BMJ* 2003; 327: 1261.

9. Studer M, Ritschard G, Gabadinho A, et al. Discrepancy Analysis of State Sequences. *Sociol Methods Res* 2011; 40: 471–510.

10. Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Serv Res* 2015; 15: 200.

11. Studer M, Ritschard G, Gabadinho A, et al. Discrepancy Analysis of Complex Objects Using Dissimilarities. In: Guillet F, Ritschard G, Zighed DA, et al. (eds) *Advances in Knowledge Discovery and Management*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3–19.

12. Bayat S, Macher MA, Couchoud C, et al. Individual and Regional Factors of Access to the Renal Transplant Waiting List in France in a Cohort of Dialyzed Patients: Renal Transplant Waitlisting in France. *Am J Transplant* 2015; 15: 1050–1060.

13. Lefort M, Vigneau C, Laurent A, et al. Facilitating access to the renal transplant waiting list does not increase the number of transplantations: comparative study of two French regions. *Clin Kidney J* 2016; 9: 849–857.

14. Couchoud C, Stengel B, Landais P, et al. The renal epidemiology and information network (REIN): a new registry for end-stage renal disease in France. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc* 2006; 21: 411–418.

15. Gabadinho A, Ritschard G, Müller NS, et al. Analyzing and Visualizing State Sequences in R with TraMineR. *J Stat Softw*; 40. Epub ahead of print 2011. DOI: 10.18637/jss.v040.i04.

16. Bergroth L, Hakonen H, Raita T. A survey of longest common subsequence algorithms. IEEE Comput. Soc, pp. 39–48.

17. Studer M, Ritschard G. What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *J R Stat Soc Ser A Stat Soc* 2016; 179: 481–511.

18. Abbott A, Forrest J. Optimal Matching Methods for Historical Sequences. *J Interdiscip Hist* 1986; 471–494.

19. Breiman L (ed). *Classification and regression trees*. Repr. Boca Raton: Chapman & Hall [u.a.], 1998.

20. Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*. Second edition, corrected at 12th printing 2017. New York: Springer, 2017.

21. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computinghttps://www.R-project.org/ (2017).

22. Studer M. WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. Epub ahead of print 2013. DOI: 10.12682/lives.2296-1658.2013.24.

23. Darak S, Mills M, Kulkarni V, et al. Trajectories of Childbearing among HIV Infected Indian Women: A Sequence Analysis Approach. *PLOS ONE* 2015; 10: e0124537.

24. Vigneau C, Kolko A, Stengel B, et al. Ten-years trends in renal replacement therapy for end-stage renal disease in mainland France: Lessons from the French Renal Epidemiology and Information Network (REIN) registry. *Néphrologie Thérapeutique*. Epub ahead of print February 2017. DOI: 10.1016/j.nephro.2016.07.453.

25. Bradbury BD, Fissell RB, Albert JM, et al. Predictors of early mortality among incident US hemodialysis patients in the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Clin J Am Soc Nephrol CJASN* 2007; 2: 89–99.

26. Rayner HC, Pisoni RL, Bommer J, et al. Mortality and hospitalization in haemodialysis patients in five European countries: results from the Dialysis Outcomes and

Practice Patterns Study (DOPPS). *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc* 2004; 19: 108–120.

27. Kim K. Discrepancy Analysis of Activity Sequences: What Explains the Complexity of People's Daily Activity-Travel Patterns? *Transp Res Rec J Transp Res Board* 2014; 2413: 24–33.

28. Moreno-Black G, Boles S, Johnson-Shelton D, et al. Exploring Categorical Body Mass Index Trajectories in Elementary School Children. *J Sch Health* 2016; 86: 495–506.

29. Drezen E, Guyet T, Happe A. From medico-administrative databases analysis to care trajectories analytics: an example with the French SNDS. *Fundam Clin Pharmacol*. Epub ahead of print 16 September 2017. DOI: 10.1111/fcp.12323.

30. MacEwan JP, Forma FM, Shafrin J, et al. Patterns of Adherence to Oral Atypical Antipsychotics Among Patients Diagnosed with Schizophrenia. *J Manag Care Spec Pharm* 2016; 22: 1349–1361.