

Identifying and Investigating Ambulatory Care Sequences Before Invasive Coronary Angiography

Anna Novelli, MA Dott. mag.,*† Julia Frank-Tewaag, MSc,† Julian Bleek, Dr. med.,‡
Christian Günster, Dipl.-Math.,§ Udo Schneider, Dr. rer. pol.,|| Ursula Marschall, Dr. med.,¶
Kathrin Schlöfler, Dr. med.,#** Norbert Donner-Banzhoff, Dr. med.,#
and Leonie Sundmacher, Dr. rer. oec.*

Background: The concept of care pathways is widely used to provide efficient, timely, and evidence-based medical care. Recently, the investigation of actual empirical patient pathways has gained attention. We demonstrate the usability of State Sequence Analysis (SSA), a data mining approach based on sequence clustering techniques, on comprehensive insurance claims data from Germany to identify empirical ambulatory care sequences. We investigate patients with coronary artery disease before invasive coronary angiography (CA) and compare identified patterns with guideline recommendations. This patient group is of particular interest due to high and regionally varying CA rates.

Methods: Events relevant for the care of coronary artery disease patients, namely physician consultations and medication prescriptions, are identified based on medical guidelines and combined to define states.

From the *Technical University of Munich; †Institute for Medical Information Processing, Biometry, and Epidemiology, Pettenkofer School of Public Health, LMU Munich; ‡Federal Association of the AOK (AOK Bundesverband); §AOK Research Institute, Berlin, Germany; ||Health Services Management, Techniker Krankenkasse, Hamburg; ¶BARMER Institut für Gesundheitssystemforschung (BARMER Institute for Health System Research), Wuppertal; #Department of General Practice and Family Medicine, University of Marburg, Marburg; and **Department of General Practice and Family Medicine, Ruhr-University Bochum, Bochum, Germany.

Data used in the analyses presented in this article are routinely collected insurance claims. Data were pseudonymized and partially reduced to impede reidentification of individuals. Approval for data linkage and data transfer to analysts was asked and granted in advance by the German Federal Social Insurance Authority (Bundesversicherungsamt) under §75 SGB X in case of nationwide statutory health insurances.

Supported by the innovation fund of the German Federal Joint Committee (Innovationsfond des Gemeinsamen Bundesausschuss) under grant number 01VSF16048.

The authors declare no conflict of interest.

Correspondence to: Anna Novelli, MA Dott. mag., Chair of Health Economics, Technical University of Munich, Munich, Georg-Brauchle-Ring 60/62, Munich 80992, Germany. E-mail: anna.novelli@tum.de.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0025-7079/22/6008-0602

State sequences are determined for 1.5 years before CA. Sequence similarity is defined for clustering, using optimal matching with theory-informed substitution costs. We visualize clusters, present descriptive statistics, and apply logistic regression to investigate the association of cluster membership with subsequent undesired care events.

Results: Five clusters are identified, the included patients differing with respect to morbidity, urbanity of residential area, and health care utilization. Clusters exhibit significant differences in the timing, structure, and extent of care before CA. When compared with guideline recommendations, 3 clusters show signs of care deficits.

Conclusions: Our analyses demonstrate the potential of SSA for exploratory health care research. We show how SSA can be used on insurance claims data to identify, visualize, and investigate care patterns and their deviations from guideline recommendations.

Key Words: patient pathway analysis, sequence clustering, ambulatory treatment pathways, coronary artery disease, insurance claims data

(*Med Care* 2022;60: 602–609)

The concept of treatment pathways has become a recognized instrument for process optimization, quality assurance, and improved efficiency of care. Pathway-based programs and treatment guidelines are being increasingly applied in the clinical and the outpatient sector to encourage treatment according to current medical evidence and to guide patients through the health system.¹ They are strongly normative in character and aim to provide an “ideal care path.” More recently, the investigation of actual, empirical patient paths has gained attention. By analyzing treatment patterns or sequences in comprehensive data, the aim is to obtain a more holistic insight into care processes and to identify possible care deficits or deviations from desired treatment paths. In addition, factors that impact on care and correlations between patterns and health-relevant outcomes are investigated.^{2–5} Data mining methods such as pattern mining, clustering, and classification methods are used to identify and investigate patient pathways, hereby exploiting the increasing availability of comprehensive data.^{3,5–7} Recently, first efforts have been made to apply a class of sequence clustering methods rooted in sociology and referred to as “State Sequence Analysis” (SSA),^{8,9} on health data to identify, visualize, and investigate empirical patient pathways and care utilization patterns.^{7,10–12}

In this study, we apply SSA to German health insurance claims to investigate the empirical ambulatory care sequences of

patients with coronary artery disease (CAD) before undergoing invasive coronary angiography (CA) using a left heart catheter. CA allows the visualization of the coronary vasculature and is used in patients with known or suspected CAD, usually in connection with and as a basis for decision-making for revascularization therapy, namely coronary artery bypass graft or percutaneous coronary intervention (PCI).^{13,14} This patient group is of particular interest: In Germany, CA rates have increased in recent decades,^{15,16} are strikingly high when compared with other countries,^{17–21} and show considerable unexplained regional variation.^{16,22} This has led to an ongoing discussion as to whether these rates reflect the actual need for CA or whether CA is performed too frequently in patients with stable CAD as a standard diagnostic procedure, or before conservative medical therapy options are exhausted.^{16,17,22,23} These deviations from national and international treatment guidelines^{13,14,24} may be driven by supply structures, medical uncertainty regarding the indication, patient preferences, and established, regionally prevailing medical practice and treatment paradigms.^{16,17,22,23,25}

The aim of the SSA presented in this study is to provide insight into the care process that precedes the CA and to compare it with guideline recommendations, in particular regarding the use of conservative management options. As a methodological contribution, we demonstrate how to jointly analyze care events of different thematic areas that can occur simultaneously, namely physician visits and medication use. For clustering, we advocate the use of a theory-driven approach to determine sequence similarity from a medical and health service perspective. Clusters are visualized and investigated using correlation analyses. In particular, we were interested in routinely performed follow-up CAs, which are strongly discouraged by current guidelines^{13,24} and can be seen as a sign of an inappropriate indication,^{22,26,27} but may be still prevalent in everyday medical practice.^{23,25}

METHODS

Data and Study Population

We conducted a longitudinal cohort study using routinely collected insurance claims data obtained from 3 German health insurances (AOK, BARMER, Techniker Krankenkasse) that share a comprehensive basic reimbursement catalogue and covered ~41% of the German population in 2016.^{28,29} The data encompasses all adults undergoing CA in the year 2016, detailing all the patients' reimbursable inpatient, outpatient and prescription claims between 2014 and 2017, diagnoses and basic demographic information. From this data, the study population was defined. All patients who underwent the index event of a CA in July or October 2016 were eligible. The available data allowed to define for each patient an observation period of 1.5 years preceding the CA and a preobservation period of 1 year preceding the observation period (Fig. 1). The restriction to July and October is due to study design reasons (see the Definition of states and sequences section). Based on clinical events recorded during preobservation and observation period, the study population was further restricted to enhance the similarity in regard to CAD morbidity: We included only patients: (1) with known CAD, that is, for whom a diagnosis of stable CAD was recorded in at least 2 quarters within the outpatient sector or at

least once within the hospital in the preobservation period; (2) who did not undergo revascularization therapy of any type or CA in the 2.5 years preceding the index CA; and (3) for whom no acute coronary event was registered during the observation period, to account for the considerable difference in clinical state that necessitate diverging treatment recommendations. Only patients, who were continuously insured between 2014 and 2017, or until death during this period, were included.

State Sequence Analysis

The SSA is conducted in 4 steps: (1) choice of care events; (2) definition of states and sequences; (3) definition of (dis-)similarity and clustering; and (4) investigation of clusters. These steps are described as follows. Analyses were performed with R and Stata, including the TraMineR package^{9,30} and the comorbidity package.³¹

Relevant Care Events

Four care events, that are reliably traceable in claims data, were chosen for consideration in the sequences:

- G: consultation of a general practitioner.
- C: consultation of an office-based cardiologist.
- P: prescription of prognosis-improving medication.
- S: prescription of symptomatic medication.

This selection was based on the national treatment guideline for stable CAD,¹³ on analyses of national and international guidelines,³² and on medical expert opinion. Prognosis-improving medication, which is recommended for every patient due to its effect on CAD morbidity and mortality,^{13,14} includes all lipid modifying agents licensed in Germany. Symptom-oriented medication, which is to be applied depending on symptom level and comorbidities of the patient, covers beta-blockers, calcium channel blockers, ivabradine, ranolazine, and organic nitrates.³² Other clinical data points, such as the medication of antiplatelet agents (obtainable without prescription), laboratory results, and lifestyle changes were not considered since these are not covered by claims data.

We derived the physician's specialty based on specialty-specific billing codes. Patients with a relevant prescription issued by a physician, whose specialty group could not be determined or was not included in the analysis, were excluded (217 patients).

Definition of States and Sequences

First, the time unit and duration of sequences, determining the observation time, were defined. Due to quarterly-based remuneration schemes in the German outpatient sector, care events as "physician consultations" can only be determined reliably within a yearly quarter. Thus, a quarter of a year is the smallest possible time unit and was used for this analysis. For each patient, the sequence period was defined as the 6 quarters before the index CA. The quarter in which the CA took place is itself not part of the observation period (Fig. 1). This necessitated the restriction to patients with the CA in the first month of a quarter (July or October 2016) to minimize the unobserved time between the end of the sequence and the index CA.

States are specified as combinations of the 4 chosen care events. Thirteen states were included, N, G, C, GC, GP,

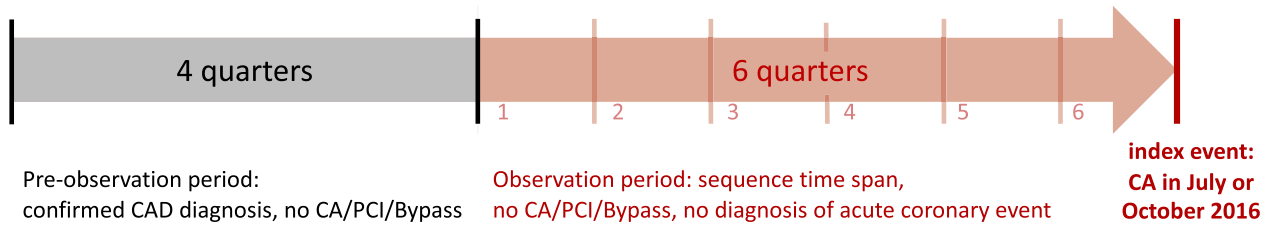


FIGURE 1. Visualization of preobservation and observation period preceding the index CA. CA indicates coronary angiography; CAD, coronary artery disease; PCI, percutaneous coronary intervention.

CP, GCP, GS, CS, GCS, GSP, CSP, GCSP with state “N” denoting the state in which no care events were recorded. For each patient and each quarter of the observation period, the patient’s state was determined by examining which of the state-defining care events had been recorded. Thus, each patient’s state sequence consisted of 6 consecutive states.

Definition of Dissimilarity and Clustering

A crucial step in clustering is the choice of a (dis-)similarity measure defining when 2 objects, or sequences, are considered similar or, conversely, dissimilar.^{30,33} We applied a localized optimal matching approach to determine sequence dissimilarity.^{30,34} Within this approach, we used Gower distance to assure that state similarity and, consequently, sequence similarity reflects the similarity of care from a health services and medical perspective.^{30,35} For clustering, we used a partitioning around medoids algorithm and performed clustering for different numbers of initial medoids (between 3 and 10), thus different numbers of clusters.³⁶ The optimal number of clusters was determined using the weighted average silhouette width.³⁷ Further details are described in the Supplemental Material (Supplemental Digital Content 1, <http://links.lww.com/MLR/C467>).

Statistical Investigation of Clusters

Clusters were visualized using frequency plots and distribution plots. We calculated summary statistics and performed χ^2 tests to investigate clusters and their unadjusted correlation with patient characteristics.

Analysis of Re-catheterization

We used logistic regression to study the discouraged medical practice of re-catheterization (CA) for controlling purposes only. We included membership to previously identified clusters as a patient characteristic into the regression model. We defined re-CA as a second invasive CA within 180 days of the index CA. To identify routinely scheduled re-CAs, we excluded patients for whom an acute coronary event was recorded concurrently with the re-CA and disregarded CAs performed in the context of the index CA (recorded within 30 d of the index CA or in the same billing case as the index CA). Patients who died within the 180-day period were excluded.

The covariables used in the regression were cluster membership, patient characteristics (age, sex),^{38,39} comorbidity indices (van Walraven-Elixhauser score),^{40,41} regional characteristics (degree of urbanization, residence in East/West Germany),⁴² variables reflecting health services utilization

[enrollment in the CAD Disease Management Program (DMP), use of noninvasive diagnostics (NID)] and variables characterizing the circumstances of the index CA (PCI, bypass, acute coronary event).^{16,17,22,26,43} Further details are provided in the Supplemental Material (Supplemental Digital Content 1, <http://links.lww.com/MLR/C467>). Regression results are reported with odds ratios and their 95% confidence intervals. A *P*-value of *P* < 0.05 was regarded as statistically significant.

RESULTS

Cluster Analysis

The study population consists of 11,535 patients. Patient characteristics are shown in Table 1.

Cluster Identification

Five clusters were identified (weighted average silhouette width = 0.35), with their size varying between 963 and 4145 patients. Figure 2 shows for each cluster the 10 most frequent sequences, the quarterly distribution of states and of events.

Each cluster is dominated by 1 or 2 states. Cluster 1 is dominated by State G (hereinafter termed “cluster G”), cluster 2 by states G and GPS (“cluster Mix”), cluster 3 by state GP (“cluster P”), cluster 4 by state GS (“cluster S”), and cluster 5 by the states GPS and GCPS (“cluster PS”).

Physician Consultations. The strong role of the general practitioner is clearly visible: Each of the cluster-dominating states includes event G. Overall rates of cardiologist involvement are comparatively low and states with event C without G are negligible. The lowest and highest cardiologist participation can be observed in cluster G and cluster Mix, respectively.

Medication Events. Cluster-dominating states differ by medication events. In clusters Mix and PS (19.9% of the study population), most patients received medication from both classes in the last 2 quarters before CA. In clusters G and S (44.2% of the study population) the absence of continuous prognosis-improving therapy before CA can be observed for most patients.

Cluster Dynamics. Across clusters, an increase in health care use towards the CA is visible. In clusters S, P, and PS, this change occurs mainly in the sixth quarter by means of a shift from the dominant state to the corresponding state with cardiologist involvement. However, this dynamic is moderate, and many patients experience a rather stable care situation throughout the observation period in these 3 clusters.

TABLE 1. Unadjusted Summary Statistics by Sequence Clusters

Patient Characteristics	Total	G	Mix	P	S	PS	P
No. patients	11,535	2261	1334	963	2832	4145	
% of study population		19.6	11.6	8.3	24.6	35.9	
Age (% of patients) (y)							
< 69	23.7	27.3	30.7	24.3	18.4	22.8	<0.001
69–76	28.5	27.5	31.4	28.4	24.5	30.8	
77–80	21.8	20.7	19.5	22.4	22.7	22.4	
> 80	26.1	24.4	18.4	24.9	34.4	24.0	
Women	37.0	33.5	25.9	30.1	47.3	37.1	<0.001
Living area (% of patients)							
Major city	25.8	27.8	29.2	25.3	24.0	25.0	<0.001
Urban area	34.9	37.6	40.9	35.9	30.6	34.1	
Rural area, densely populated	19.1	16.6	16.0	19.4	20.9	20.1	
Rural area, sparsely populated	20.3	18.0	13.9	19.3	24.5	20.8	
Region: patients living in East Germany	25.8	23.2	22.3	23.1	31.1	25.2	<0.001
Patients with an acute coronary event within CA billing case	24.1	24.4	23.2	24.5	25.1	23.4	0.435
Elixhauser Comorbidity Score							
Based on ambulatory diagnoses							
Mean (SD)	10.8 (9.3)	9.4 (9.2)	10.2 (9.3)	10.3 (9.1)	11.5 (9.4)	11.3 (9.3)	<0.001
Based on hospital diagnoses							
Mean (SD)	10.0 (9.3)	9.0 (9.0)	8.8 (9.1)	8.6 (9.2)	11.5 (9.5)	10.2 (9.3)	<0.001
Patients enrolled in DMP	48.3	35.6	53.5	54.2	41.9	56.5	<0.001
Patients with noninvasive diagnostics within 3 mo before CA	33.8	33.4	41.9	35.7	28.9	34.3	<0.001
Patients receiving invasive procedure following CA within 30 d after index CA or within same billing case							
PCI	38.6	38.6	39.9	38.3	40.3	37.0	0.067
CABG	4.8	5.1	5.3	4.2	4.4	4.8	0.534
Patients with re-CA within 30–180 d after index CA*	6.4	6.7	5.2	6.8	6.0	6.8	0.296
Patients who died within 180 d after index CA	8.6	7.5	5.6	7.2	11.6	8.3	<0.001

*For the re-CA rate, the denominator population is not the entire study population of 11,535 patients, but only 10,427 patients, since 986 patients were excluded since they died in the timeframe of 180 days and 122 patients were excluded because an acute coronary event was coded in the billing episode of the second CA.

CA indicates (invasive) coronary angiography; CABG, coronary artery bypass graft surgery; CAD, coronary artery disease; DMP, structured disease management program for coronary artery disease patients; PCI, percutaneous coronary intervention; re-CA, re-catheterization.

In contrast, cluster G and, to a considerably greater extent, cluster Mix exhibit escalating dynamics from the middle of the observation period on. Both clusters show an increase in medication events. While the increase in cluster G is small, cluster Mix reaches medication levels as high as those of cluster PS towards the CA. In addition, cluster Mix shows a gradual increase of cardiologist involvement throughout the observation period.

Patient Characteristics by Sequence Clusters

Patient characteristics differ significantly between clusters (Table 1).

Cluster Mix has the youngest, mostly urban population and the lowest percentage of women. Its patients seem to be healthier in general, with a low comorbidity index and the lowest 180-day mortality rate. Additional indicators of health care use, NID use and DMP participation, are by far the highest, respectively above average in this cluster.

Cluster S provides the greatest contrast to cluster Mix. It is characterized by a significantly older, mostly rural population, with more comorbidities, the highest mortality rate and the highest proportion of women. NID use and DMP participation were substantially lower than in other clusters.

While clusters P and PS show similarities to cluster S in terms of age distribution and the region of residence, relatively high rates of DMP enrollment and NID use can be observed.

Finally, cluster G is comparable to cluster Mix in terms of age, sex, residential area and comorbidity, but shows significantly lower rates of DMP enrollment and NID use.

Neither the occurrence of an acute coronary event, nor subsequent re-CA within 180 days, nor bypass surgery within 30 days of the CA show associations with cluster membership in unadjusted correlation analysis. A slight tendency for correlation ($P=0.07$) might be present for PCI within 30 days of CA.

Occurrence of Re-catheterization

The identified clusters were included as a predictor in a logistic regression model, with the occurrence of a second CA within 180 days (re-CA) of the index CA as dependent variable. The patient population for the regression was reduced to 10,427 patients, excluding 986 patients who died in the relevant period and 122 patients for whom an acute coronary event was diagnosed within the billing episode of the second CA. Regression results are presented in Table 2.

Across clusters, 6.4% of patients experienced a re-CA. The logistic regression reveals that by adjusting for patient characteristics and morbidity, cluster membership is significantly associated with re-CA: Patients of clusters G, P, S, and PS have higher odds compared with those of cluster Mix. Older patients show decreased odds for a re-CA. Being female is negatively associated with receiving re-CA. Significantly increased odds for re-CA are seen in patients who live in former East Germany. Eight-fold odds are seen in patients who received a PCI within 30 days of the index CA.

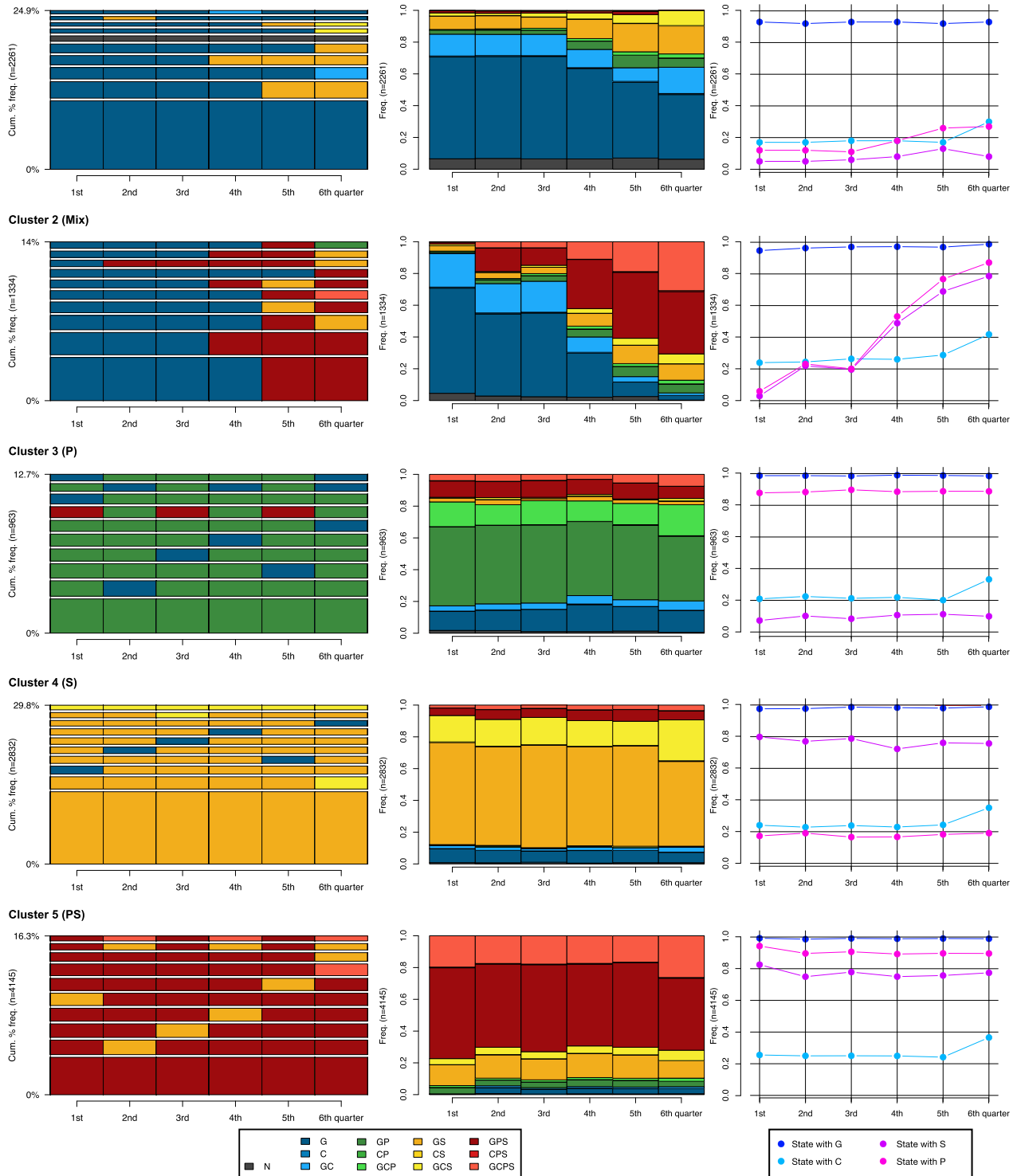


FIGURE 2. Frequency (left) and distribution plots (middle and right) for each of the 5 identified clusters. The frequency plots show the 10 most frequent sequences of each cluster. The y-axis shows cumulative frequency; thus, the height of the sequences is relative to their occurrence. The distribution plot in the middle show the distribution of states in each quarter of the observation period. The 13 states in the figures on the left and in the middle are color-coded, with the ground color (blue/green/yellow/red) used to indicate the combination of medication events (none/P/S/PS) and the brightness (dark/middle/bright) used to indicate the accompanying physician events (G/C/GC). The distribution plot on the right-hand side visualizes the event distribution throughout the observation period.

TABLE 2. Results of Logistic Regression for Re-catheterization Within 180 Days After the Index CA

Independent Variables	OR	P	95% CI
Reference: cluster Mix			
Cluster G	1.48	0.015	1.08; 2.03
Cluster P	1.51	0.031	1.04; 2.21
Cluster S	1.40	0.035	1.02; 1.92
Cluster PS	1.55	0.003	1.16; 2.07
Reference: age < 69 (y)			
69–76	0.79	0.036	0.64; 0.98
77–80	0.82	0.104	0.65; 1.04
> 80	0.54	<0.001	0.42; 0.70
Reference: male			
Female	0.70	<0.001	0.58; 0.85
Living area, reference: major city			
Urban area	1.18	0.150	0.94; 1.46
Rural, densely populated	0.93	0.576	0.73; 1.19
Rural, sparsely populated	0.98	0.871	0.76; 1.25
Region, reference: West Germany			
East Germany	1.37	0.002	1.12; 1.68
Elixhauser Comorbidity Score			
Ambulatory	1.00	0.561	0.99; 1.01
Hospital	1.00	0.431	0.99; 1.01
Revascularization procedure within 30 d, reference: none			
PCI	8.56	<0.001	6.92; 10.59
CABG	0.50	0.132	0.20; 1.23
Acute coronary event in billing case of index CA			
NID within 3 mo before CA	1.13	0.190	0.94; 1.35
DMP	1.01	0.934	0.85; 1.19
Constant	0.02	<0.001	0.01; 0.02

CA indicates (invasive) coronary angiography; CABG, coronary artery bypass graft surgery; CAD, coronary artery disease; CI, confidence interval; DMP, structured disease management program for coronary artery disease patients; NID, noninvasive diagnostic procedure; OR, odds ratio; PCI, percutaneous coronary intervention.

DISCUSSION

In this study, we apply SSA to comprehensive insurance claims data from 3 German health insurances to explore ambulatory care patterns of patients with CAD before undergoing CA. Due to the high number of observations and since the data cover multiple years, it was possible to create a study population that is approximately homogeneous in terms of CAD morbidity, despite the absence of clinical or lifestyle information.

Identified Treatment Patterns

For patients with stable CAD, our analysis demonstrates the heterogeneity of treatment patterns preceding CA with respect to physician consultations and medication. We identified and visualized 5 clusters of care sequences.

In all clusters, the majority of patients consulted general practitioners throughout the observation period. This is consistent with guideline recommendations that emphasize the responsibility of general practitioners to coordinate continuous evidence-based care.¹³ Although overall cardiologist involvement is relatively low, an increase is visible towards the index CA as recommended in the guidelines.¹³

Clusters differ with respect to the medication prescribed, the extent of cardiologist participation, and the development over time.

Two clusters, clusters Mix and PS, show a high intensity of care before CA, alongside high rates of DMP enrollment and

NID. While the older and more comorbid population of cluster PS received this level of care continuously, the younger, urban and healthier patients of cluster Mix evolve to reach this high level of care. Its patients experience treatment closely reflecting the guidelines, with optimal medication therapy, cardiologist involvement, DMP enrollment and NID use. This could be guided by the physician, by the behavior of informed patients, or by supply structures in urban areas.

The remaining 3 clusters lack the presence of 1 or both medication events: A substantial proportion of clusters P and G does not receive continuous symptom-improving medication. One might hypothesize that this absence of medication reflects the absence of symptoms. This would imply that CA was not performed out of symptomatic, but out of prognostic indication. However, the absence of higher coronary artery bypass graft rates in these clusters does not support this hypothesis. In another scenario, patients might undergo CA considering revascularization out of symptomatic indication. In this case, contrary to guideline recommendations, revascularization therapy seems to have been considered before conservative symptomatic drug therapy options had been exhausted.

Patients from clusters G and S do not receive continuous prognosis-improving medication, even immediately preceding CA. This indicates deficits in the care of stable CAD patients and a deviation from guideline recommendations. Notably, cluster S has the highest proportion of female patients. This reminisces the repeatedly stated hypothesis of underestimation and undertreatment of CAD in women.^{38,39,44–46}

To gain additional insight, a logistic regression was performed with re-CA within 180 days as the outcome. The observed re-CA rate of 6.4% is substantial, yet significantly lower than in a previous study,²⁵ most likely due to our conservative approach in the operationalization of re-CA.

The regression revealed that older and more comorbid patients are less likely to receive a re-CA. This meets our expectations since providers might refrain from routinely scheduling of re-CAs for patients with higher procedural risks. A similar reasoning was hypothesized by Piedmont et al²² in a comparable context. Our results on the correlation between PCI at index event and re-CA also complement the descriptive results of Jeschke et al.²⁵ With respect to cluster membership, we find that patients of cluster Mix have lower odds for re-CA compared with all other clusters. This confirms the characterization of cluster Mix as being in good accordance with guidelines. A possible explanation could be that for patients from this cluster, being the youngest and healthiest, a second intervention, even only for controlling purposes, seems unnecessary. Another probable scenario is that these patients or their physicians refrain to a higher degree from performing re-CA in general.

Strengths and Limitations

This is the first SSA on insurance claims data that includes different kinds of simultaneous or overlapping care events, namely physician visits and medication, in a single analysis. Previous studies have focused on the frequency or volume of health service utilization alone, or on different care events in separate analysis.^{7,10–12} Our combined approach enables to exploit the potential of the SSA method to capture

the treatment of large patient population in a holistic manner over a considerable period. The method is however limited to a rather small number of care events per type (due to the growing number of resulting states). This also implies that health care can be captured only with respect to the selected events. We identified 5 clusters. The silhouette coefficient, which reflects intracluster homogeneity and intercluster heterogeneity, is 0.35. Comparing to previous studies,^{7,10,11} this value indicates a high level of cluster quality, in particular since our combined approach is done at the expense of a larger number of states with a corresponding increase in sequence heterogeneity. We advocate the use of a theory-driven definition of state and consequently sequence dissimilarity. This allows to capture the medical relevance of a transition between 2 states, which is not guaranteed for standard data-driven dissimilarity measures.

From a health care management perspective, our results provide health policymakers and care providers with easily accessible and comprehensive insights into the health care situation of CAD patients in Germany. The identified deviations from guidelines may indicate aims of targeted interventions.

Since the study population consists of patients with diagnosed stable CAD, the study results only relate to stable CAD patients. Despite their comprehensiveness, insurance claims data have some limitations. The reasons for physician consultations are not reported and could also be noncardiac issues. Since physicians in Germany receive remuneration partly in quarterly flat rates, physician visits can only be determined on a quarterly basis, and the frequency of visits per quarter cannot be reliably determined. Further, the data do not contain clinical or lifestyle information. Thus, it is not possible to determine whether the indication for CA reflects morbidity-based need or observed deviations from guideline care do indicate suboptimal therapy. For example, a fast progression of symptoms may necessitate early revascularization. However, it can be reasonably assumed that such phenomena occur at a much smaller scale than the macrolevel picture enabled by SSA. For instance, our analysis revealed that 2 clusters, constituting substantial 44% of the study population, do not receive continuing prognosis-improving medication before CA.

CONCLUSIONS

This study investigated ambulatory care sequences preceding invasive CA among patients with stable CAD. To this end, we applied sequence clustering techniques on German health insurance claims. Different types of care events were considered within a combined SSA. Based on a theory-driven approach to determine sequence similarities, 5 clusters of treatment patterns were identified. Alongside CAD morbidity and comorbidity, regional structures, patient preferences, or medical practice patterns might influence treatment pathways. The comparison to guideline recommendations suggests the presence of care deficits within some clusters and may indicate starting points of further research and targeted interventions. Logistic regression revealed that cluster membership is correlated with the risk of subsequent health interventions. Future studies should further explore the potential of SSA for patient phenotyping, risk stratification, and predictive modeling of health outcomes and health care use.

REFERENCES

- Schrijvers G, van Hoom A, Huiskes N. The care pathway: concepts and theories: an introduction. *Int J Integr Care*. 2012;12:e192.
- Hanson CL, Osberg M, Brown J, et al. Conducting patient-pathway analysis to inform programming of tuberculosis services: methods. *J Infect Dis*. 2017;216:S679–S685.
- Huang Z, Lu X, Duan H. On mining clinical pathway patterns from medical behaviors. *Artif Intell Med*. 2012;56:35–50.
- Li X, Mei J, Liu H, et al. Analysis of care pathway variation patterns in patient records. *Stud Health Technol Inform*. 2015;210:692–696.
- Lakshmanan GT, Rozsnyai S, Wang F. Investigating clinical care pathways correlated with outcomes. In: Daniel F, Wang J, Weber B, eds. *Business Process Management: Lecture Notes in Computer Science*. Berlin/Heidelberg, Germany: Springer; 2013:323–338.
- Rojas E, Munoz-Gama J, Sepúlveda M, et al. Process mining in healthcare: a literature review. *J Biomed Inform*. 2016;61:224–236.
- Vogt V, Scholz SM, Sundmacher L. Applying sequence clustering techniques to explore practice-based ambulatory care pathways in insurance claims data. *Eur J Public Health*. 2018;28:214–219.
- Abbott A, Tsay A. Sequence analysis and optimal matching methods in sociology. *Sociol Methods Res*. 2000;29:3–33.
- Gabardino A, Ritschard G, Müller NS, et al. Analyzing and visualizing state sequences in R with TraMineR. *J Stat Softw*. 2011;40:1–37.
- Roux J, Grimaud O, Leray E. Use of state sequence analysis for care pathway analysis: the example of multiple sclerosis. *Stat Methods Med Res*. 2019;28:1651–1663.
- 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;15:1–10.
- Le Meur N, Vigneau C, Lefort M, et al. Categorical state sequence analysis and regression tree to identify determinants of care trajectory in chronic disease: Example of end-stage renal disease. *Stat Methods Med Res*. 2019;28:1731–1740.
- Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV), and Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). Nationale Versorgungsleitlinie Chronische KHK [National treatment guideline for chronic coronary heart disease]. Auflage. 5; Version 1; 2019.
- Montalescot G, Sechtem U, Achenbach S, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the Management of Stable Coronary Artery Disease of the European Society of Cardiology. *Eur Heart J*. 2013;34:2949–3003.
- Statistisches Bundesamt. Operationen und Prozeduren der vollstationären Patientinnen und Patienten in Krankenhäusern (Wohnort/Behandlungsort) (Primärquelle: DRG-Statistik und PEPP-Statistik, Statistisches Bundesamt) [In patient hospital surgeries and procedures (place of residence/place of treatment)]; 2019.
- Möckel M, Searle J, Jeschke E. Indikation, Prognose und regionale Unterschiede der Herzkatheterversorgung in Deutschland [Indication, prognosis and regional differences in cardiac catheter care in Germany]. In: Klauber J, Günster C, Gerste B, Robra BP, Schmacke N, eds. *Versorgungs-Report 2013/2014 [Healthcare report 2013/2014]*. Stuttgart, Germany: Schattauer; 2014:231–254.
- Dissmann W, Ridder MD. The soft science of German cardiology. *Lancet*. 2002;359:2027–2029.
- Organisation for Economic Co-operation and Development (OECD). *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance? OECD Health Policy Studies*. Paris, France: OECD Publishing; 2014.
- Jennings S, Bennett K, Shelley E, et al. Trends in percutaneous coronary intervention and angiography in Ireland, 2004–2011: implications for Ireland and Europe. *Int J Cardiol Heart Vessel*. 2014;4:35–39.
- Cook S, Walker A, Hugli O, et al. Percutaneous coronary interventions in Europe: prevalence, numerical estimates, and projections based on data up to 2004. *Clin Res Cardiol*. 2007;96:375–382.
- Barbato E, Noc M, Baumbach A, et al. Mapping interventional cardiology in Europe: the European Association of Percutaneous Cardiovascular Interventions (EAPCI) Atlas Project. *Eur Heart J*. 2020;41:2579–2588.
- Piedmont S, Swart E, Kenmogne R, et al. Linksherzkatheteruntersuchungen und ihre invasive Konsequenz—Regionalvergleiche ermitteln

- auffällige Unterschiede [Left-heart catheterization followed by other invasive procedures: Regional comparisons reveal peculiar differences]. *Z Evid Fortbild Qual Gesundheitswes.* 2017;127-128:62–71.
23. Herwig A, Dehnen D, Weltermann B. Patient factors driving overuse of cardiac catheterisation: a qualitative study with 25 participants from two German teaching practices. *BMJ Open.* 2019;9:e024600.
 24. Hamm CW, Albrecht A, Bonzel T, et al. Diagnostische Herzkatheteruntersuchung [Diagnostic cardiac catheterization]. *Clin Res Cardiol.* 2008;97:475–512.
 25. Jeschke E, Baberg HT, Dirschedl P, et al. Komplikationen und Folgeeingriffe nach koronaren Prozeduren in der klinischen Routine [Complication rates and secondary interventions after coronary procedures in clinical routine]. *Dtsch Med Wochenschr.* 2013;138:570–575.
 26. Jeschke E, Günster C. Qualitätsindikatoren für Koronarangiographie: Sektorenübergreifende Qualitätsmessung auf der Basis von Routinedaten [Quality indicators for coronary angiography: Cross-sectoral quality measurement based on routine data]. In: Klauber J, Geraedts M, Friedrich J, Wasem J, eds. *Krankenhaus-Report 2013 [Hospital-Report 2013]—Mengendynamik: mehr Menge, mehr Nutzen?* Stuttgart, Germany: Schattauer; 2013:263–279.
 27. Baldus S, Werdan K, Levenson B, et al. Klug entscheiden: ...in der Kardiologie [Decide wisely: ...in Cardiology]. *Dtsch Arztl.* 2016;113:1312–1315.
 28. Statistisches Bundesamt. Indikator 1 der ECHI shortlist: Bevölkerung nach Geschlecht und Alter (Primärquelle: Fortschreibung des Bevölkerungsstandes, Statistisches Bundesamt) [Indicator 1 of the ECHI shortlist: Population by gender and age (primary source: update of population status, Statistisches Bundesamt)]; 2016.
 29. Bundesministerium für Gesundheit. Gesetzliche Krankenversicherung—Mitglieder, mitversicherte Angehörige und Krankenstand—Jahresdurchschnitt 2016 [Statutory health insurance Members, co-insured dependents and sick leave—annual average 2016]; 2016.
 30. 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.
 31. Gasparini A. Comorbidity: an R package for computing comorbidity scores. *J Open Source Softw.* 2018;3:648.
 32. Frank-Tewaag J, Bleek J, Horenkamp-Sonntag D, et al. Use of guideline-recommended drug therapy in patients undergoing percutaneous coronary intervention for stable coronary heart disease in Germany: a multilevel analysis of nationwide routine data. *BMJ Open.* 2020;10:e042886.
 33. Lesnard L. Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociol Methods Res.* 2010;38:389–419.
 34. Hollister M. Is optimal matching suboptimal? *Sociol Methods Res.* 2009;38:235–264.
 35. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics.* 1971;27:857.
 36. Kaufman L, Rousseeuw PJ. Partitioning Around Medoids (Program PAM). *Finding Groups in Data.* Hoboken, New Jersey: John Wiley Sons Inc.; 1990:68–125. Available at: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316801>.
 37. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987;20:53–65.
 38. Shaw LJ, Bairey Merz CN, Pepine CJ, et al. Insights from the NHLBI-Sponsored Women's Ischemia Syndrome Evaluation (WISE) Study: part I: gender differences in traditional and novel risk factors, symptom evaluation, and gender-optimized diagnostic strategies. *J Am Coll Cardiol.* 2006;47:S4–S20.
 39. Daly C, Clemens F, Lopez Sendon JL, et al. Gender differences in the management and clinical outcome of stable angina. *Circulation.* 2006;113:490–498.
 40. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care.* 1998;36:8–27.
 41. van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care.* 2009;47:626–633.
 42. Lampert T, Müters S, Kuntz B, et al. Beschreibung der gesundheitlichen Lage in Ost und West sowie deren Entwicklung in den letzten 30 Jahren [30 years after the fall of the Berlin Wall: Regional health differences in Germany]. *J Health Monit.* 2019;4:2–25.
 43. Inoue T, Kuwabara H, Fushimi K. Regional variation in the use of percutaneous coronary intervention in Japan. *Circ J.* 2017;81:195–198.
 44. Virani SS, Woodard LD, Ramsey DJ, et al. Gender disparities in evidence-based statin therapy in patients with cardiovascular disease. *Am J Cardiol.* 2015;115:21–26.
 45. Victor BM, Teal V, Ahedor L, et al. Gender differences in achieving optimal lipid goals in patients with coronary artery disease. *Am J Cardiol.* 2014;113:1611–1615.
 46. Koopman C, Vaartjes I, Heintjes EM, et al. Persisting gender differences and attenuating age differences in cardiovascular drug use for prevention and treatment of coronary heart disease, 1998–2010. *Eur Heart J.* 2013;34:3198–3205.