Cluster analysis and its utility in healthcare claims data analysis: results of a retrospective observational study of patients with metastatic urothelial cancer in Germany

MSR157

M. Kearney,[1] F. Hardtstock,[2] J. Krieger,[2*] B. Deiters,[3] U. Maywald,[4*] U. Osowski,[5] T. Wilke,[6,7] G. Niegisch,[8] M.-O. Grimm[9]

[1]Merck Healthcare KGaA, Darmstadt, Germany; [2]Cytel, Berlin, Germany; [3]GWQ ServicePlus AG, Düsseldorf, Germany; [4]AOK PLUS, Dresden, Germany; [5]Merck Healthcare Germany GmbH, Weiterstadt, Germany, an affiliate of Merck KGaA; [6]IPAM e.V., Wismar, Germany; [7]GIPAM, Wismar, Germany; [8]University Hospital and Medical Faculty of the Heinrich-Heine-University, Düsseldorf, Germany; [9]Universitätsklinikum Jena, Jena, Germany
*Affiliation at the time the study was conducted.

CONCLUSIONS

- **This is the first study of healthcare claims data using a clustering method to identify meaningful subgroups of patients with metastatic urothelial cancer (mUC) who did not receive systemic anticancer treatment**
- **A total of 1,892 patients (58.6%) with mUC did not receive systemic treatment within 12 months of diagnosis**
- **Clusters with the highest proportion of untreated patients had the largest number of patients who were older, had higher Charlson Comorbidity Index (CCI) scores, required home care, and were more likely to receive their index diagnosis in smaller hospitals**
- **The results corroborate findings from previous studies in which older patients with mUC who had several comorbidities were most likely to remain untreated[1]**
- **By employing clustering analytic techniques to identify distinct subgroups at increased risk of nontreatment, including those diagnosed in smaller hospitals, healthcare decision-makers may design personalized intervention programs according to the unique needs and circumstances of each patient**

PLAIN LANGUAGE SUMMARY

- **In this study, we used 2 claims databases (AOK PLUS and GWQ) to look at medical records of people who were receiving treatment for advanced urothelial cancer in Germany**
- **Of the 3,226 people diagnosed with advanced urothelial cancer, the average age was 73.8 years and 70.8% were men**
- **Only 1,286 people (39.9%) received drug treatment, while 1,892 people (58.6%) did not**
- **Statistical modeling showed that people who received drug treatment were likely to be younger and healthier**
- **A technique called cluster analysis was used that grouped similar people together into clusters; this helped identify patterns in their care**
- **The clusters with the most untreated people also had a larger number of older people**
  - **Untreated people were also more likely to have more comorbidities (other medical conditions), need home care, and have been diagnosed in smaller hospitals**
- **More studies are needed to investigate the reasons why some eligible people with advanced urothelial cancer do not receive drug treatment; this will ensure that more people can benefit from available treatments**

BACKGROUND

- The treatment landscape in mUC is rapidly evolving, with novel therapeutic advances being incorporated in guidelines and clinical practice[2-4]
- Understanding real-world treatment patterns and patient outcomes is crucial for improving care and developing treatment guidelines for mUC
- Cluster analysis (CA) is a frequently used applied statistical technique that reveals hidden structures and homogeneous clusters or subgroups in large datasets[5-7]
  - This is an exploratory approach that reveals natural grouping and aims to identify groups that are inherently very homogeneous and as heterogeneous as possible to other groups
- The purpose of this retrospective, observational cohort study was to use CA to identify clinically relevant segments of patients with mUC in Germany who did not receive systemic anticancer treatment

METHODS

Data source and study design

- This nonexperimental, retrospective CA used anonymized data from 2 statutory health insurance claims databases (2013-2020; ≈8 million insured) to identify adult patients with an incident mUC diagnosis using *ICD-10* codes C65-68 and C77-79 from 2015-2019 in Germany[8,9]
- Patients with other malignant tumors were excluded
- Patients were observed for ≥12 months after incident mUC diagnosis (index) or until death

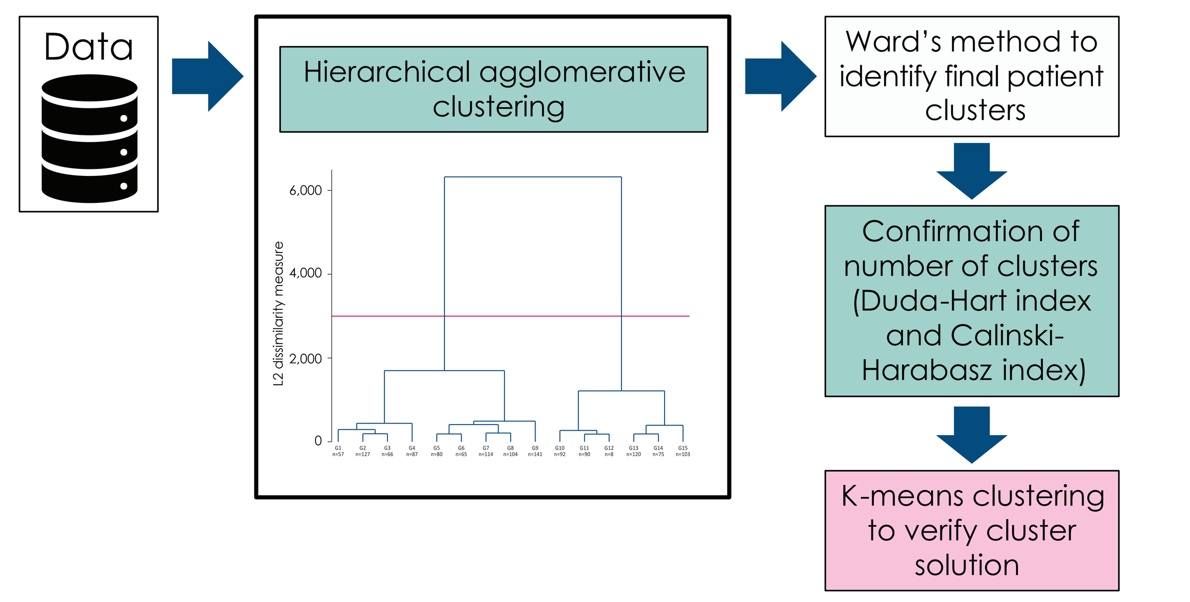
Statistical methods

- Patients' characteristics were analyzed using descriptive statistics, including absolute and relative frequencies for categorical variables and summary statistics (mean, SD) for continuous variables
- Analyses were performed separately for each database, with results first combined using meta-analysis methods and then presented in aggregate form

Cluster analysis technique

- A CA application was used to identify patient characteristics and healthcare system factors related to nontreatment (**Figure 1**)
- After checking the multicollinearity of available variables, their number was reduced so that only 1 expression of each variable with high correlation (r>0.7) was left in the dataset (eg, only 1 hospitalization variable)
- Hierarchical agglomerative clustering was performed
  - The algorithm starts by considering each observation as a single cluster, followed by merging pairs of clusters one by one until all observations or in-between clusters have been merged into 1 overall cluster. The result of the hierarchical agglomerative clustering is a dendrogram (tree-like depiction of the sequences of merges). Patients are homogeneous in a cluster and heterogeneous compared with patients assigned to other clusters
- First, outliers in single linkage were identified by looking for clusters with limited data points
- Next, the Ward's method, a hierarchical clustering algorithm that aims to minimize the total within-cluster variance when merging clusters, was conducted with continuous and binary variables to identify the final patient clusters (homogeneous in a cluster and heterogeneous compared with patients assigned to other clusters)
- The Duda-Hart Je(2)/Je(1) index, associated pseudo T-squared value, and Calinski-Harabasz index were used to confirm the number of clusters
- K-means clustering was used to verify the cluster solution
- Finally, patient characteristics were illustrated for each cluster in dichotomized (binary) form, using mean values as the cutoff for continuous data

Figure 1. Cluster analysis method

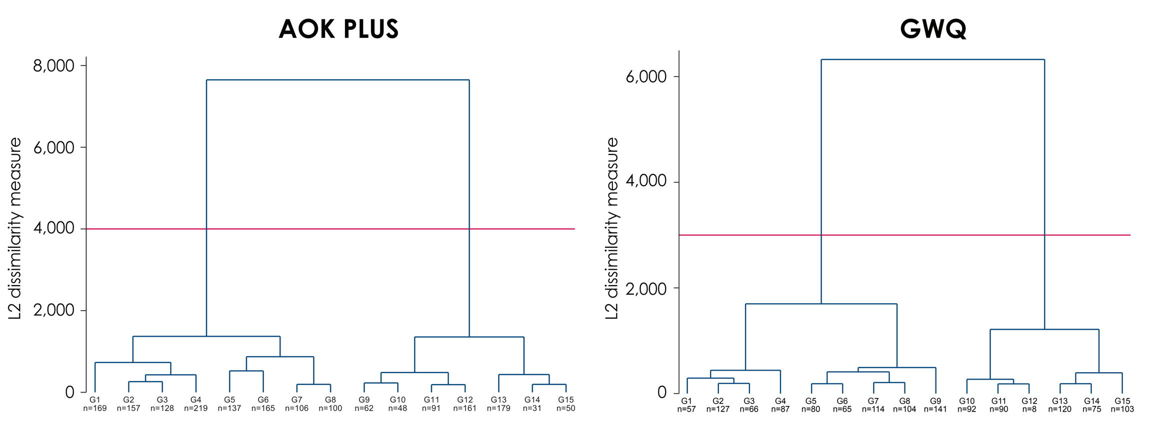


G, group; **L2**, Euclidean distance.

RESULTS

- Of 3,226 patients with mUC, 70.8% were male, mean (SD) age was 73.8 (10.8) years, mean (SD) CCI score was 6.3 (3.8), and mean (SD) Elixhauser Comorbidity Index score was 17.6 (11.4)
- A total of 1,892 patients (58.6%) with mUC did not receive systemic treatment within 12 months of diagnosis
- CA was performed, and after identifying outliers during single linkage (AOK, n=22; GWQ, n=24), Ward's linkage with untransformed variables was conducted for each database to identify patient clusters (homogeneous in a cluster and heterogeneous compared with patients assigned to other clusters)
- After identifying outliers, CA indicated that a 2-cluster solution was the most appropriate option for both databases (**Figure 2**)
- The identified patient characteristics between the 2 clusters were compared (**Table 1**), and the 2-cluster solution was visualized by using dichotomized variables (**Figure 3**)
- Clusters with the highest proportion of untreated patients (AOK PLUS, cluster 2 with 87.2% untreated patients; GWQ, cluster 1 with 59.0% untreated patients) also had the highest proportion of patients who were older (AOK PLUS, 81.9% aged ≥75 years; GWQ, 70.6% aged ≥73 years) and had higher CCI scores (AOK PLUS, 63.5% with score of ≥7; GWQ, 64.6% with score of ≥6)
- In AOK PLUS, all patients in cluster 2 required home care and were more likely to receive their index diagnosis in smaller hospitals (bed count <500) vs patients in cluster 1 (39.1% vs 28.6%, respectively; observed in k-means clustering)
- Cluster analyses for comorbidities and procedures are presented in **Supplementary Table 1** and **Supplementary Table 2**

Figure 2. Cluster analysis: dendrogram for Ward's linkage



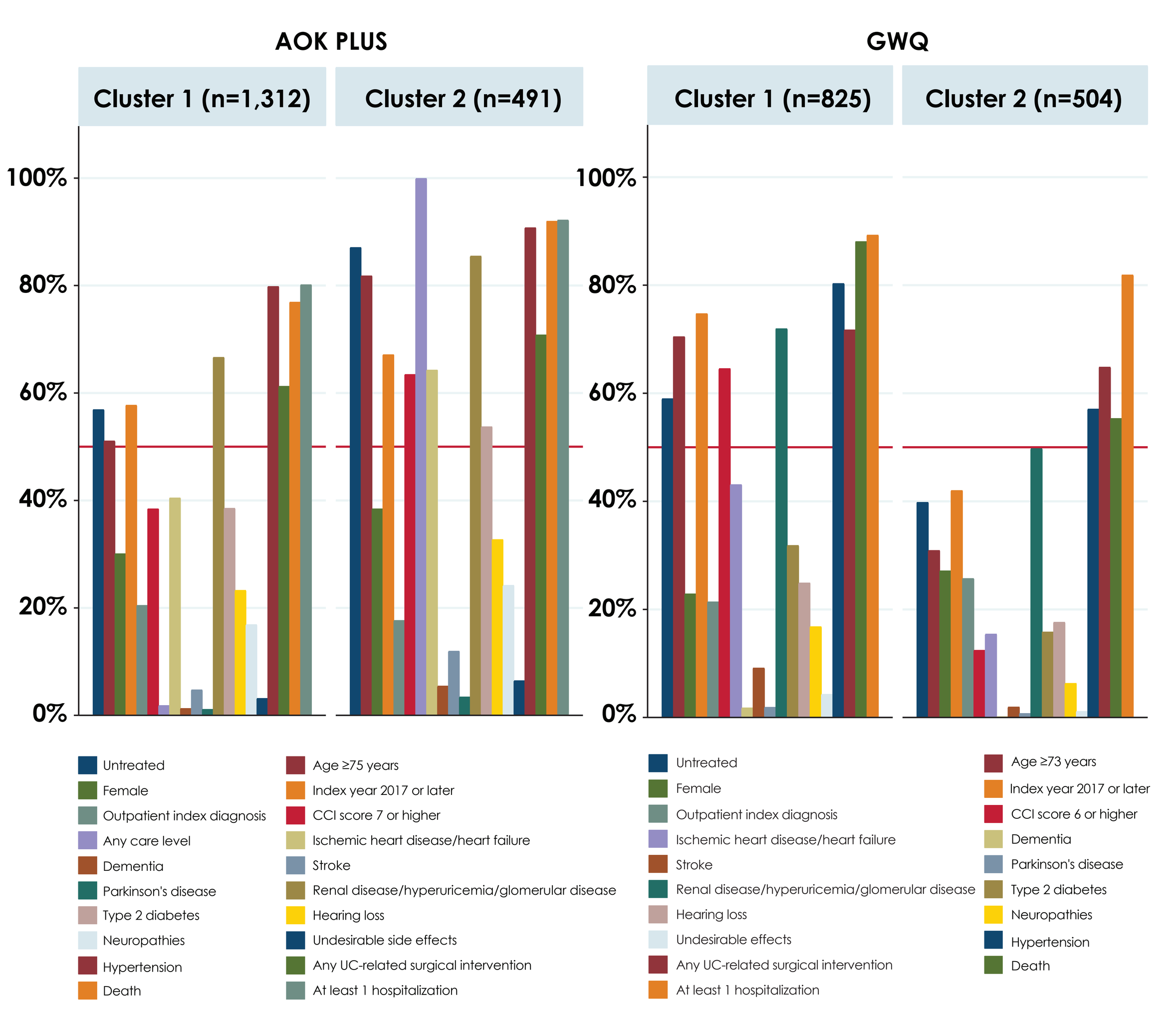
G, group; **L2**, Euclidean distance.

Table 1. Cluster analysis: 2-cluster solution using k-means

	AOK PLUS			GWQ		
	Cluster 1 (n=1,312)	Cluster 2 (n=491)	p value	Cluster 1 (n=825)	Cluster 2 (n=504)	p value
Untreated patients, n (%)	**746 (56.9)**	**428 (87.2)**	**<0.001**	**487 (59.0)**	**201 (39.9)**	**<0.001**
Age, mean (SD), years	72.5 (10.1)	80.5 (8.8)	<0.001	75.9 (10.2)	67.2 (11.0)	<0.001
Female, n (%)	393 (30.0)	189 (38.5)	0.001	189 (22.9)	137 (27.2)	0.079
Deaths, n (%)	1,009 (76.9)	452 (92.1)	<0.001	727 (88.1)	279 (55.4)	<0.001
CCI score, mean (SD)	5.9 (3.6)	8.4 (3.8)	<0.001	7.3 (3.6)	3.5 (2.2)	<0.001
ECI score, mean (SD)	16.5 (11.1)	23.7 (11.1)	<0.001	20.0 (11.2)	10.4 (8.1)	<0.001
Index year, n (%)						
2015	282 (21.5)	83 (16.9)	0.006	99 (12.0)	120 (23.8)	<0.001
2016	272 (20.7)	78 (15.9)		109 (13.2)	172 (34.1)	
2017	265 (20.2)	107 (21.8)		185 (22.4)	113 (22.4)	
2018	234 (17.8)	100 (20.4)		172 (20.9)	66 (13.1)	
2019	259 (19.7)	123 (25.1)		260 (31.5)	33 (6.6)	
Level of care, n (%)						
None	1,302 (99.2)	0	<0.001	NA	NA	–
Low/medium	10 (0.8)	402 (81.9)		NA	NA	
High	0	89 (18.1)		NA	NA	
Outpatient index mUC diagnosis, n (%)	271 (20.7)	87 (17.7)	0.164	177 (21.5)	130 (25.8)	0.069
Size of diagnosing hospital, n (%)						
<500 beds	375 (28.6)	192 (39.1)	<0.001	NA	NA	–
≥500 beds	436 (33.2)	145 (29.5)	0.135	NA	NA	–
≥1 all-cause hospitalization, n (%)	1,049 (80.0)	453 (92.3)	<0.001	737 (89.3)	413 (81.9)	<0.001
No. of all-cause hospitalizations, mean (SD)	2.6 (2.7)	4.1 (3.4)	<0.001	4.2 (4.0)	2.7 (2.8)	<0.001
Duration of all-cause hospitalization, mean (SD), days	20.4 (23.4)	41.5 (47.8)	<0.001	35.0 (43.4)	19.9 (34.4)	<0.001

CCI, Charlson Comorbidity Index; **ECI**, Elixhauser Comorbidity Index; **mUC**, metastatic urothelial cancer; **NA**, not available (in the database).

Figure 3. Cluster analysis with dichotomized variables



CCI, Charlson Comorbidity Index; **UC**, urothelial cancer.

LIMITATIONS

- The present study has several limitations. Firstly, it is retrospective and therefore subject to limitations of this type of study
- Administrative claims data are not collected for research purposes, and measurement error may have been introduced by erroneous coding or coding that was more driven by reimbursement needs than research needs
- The study inclusion period was limited to 2015-2019; since then, the approval of novel therapeutics for mUC may have led to an increase in systemic treatment rates
- Additionally, treatment rates may have been underestimated, as patients treated in investigational clinical studies appear as untreated in German claims data
- While both the AOK PLUS and GWQ datasets contain information from routine medical practice, they may not capture all relevant aspects of a patient's medical history; thus, the accuracy of cluster assignments may have been impacted, and important associations may have been overlooked
- This analysis could only address variables that were available in the datasets; for example, other variables driving treatment/nontreatment decisions such as Eastern Cooperative Oncology Group performance status or patients' wishes could not be considered
- The optimal number of clusters can be subjective; different clustering methods or criteria may yield varying cluster solutions, making it challenging to determine the correct number of clusters

GET POSTER PDF

Correspondence: **Mairead Kearney, mairead.kearney@merckgroup.com**

GET POSTER SUPPLEMENTARY DATA

Supplementary data can be accessed through this QR code

REFERENCES 1. Kearney M, et al. Future Oncol. 2023;20(16):1123-37. **2.** Flaig TW. J Natl Compr Canc Netw. 2018;16(5S):636-8. **3.** Powles T, et al. Ann Oncol. 2024;35(6):485-90. **4.** Witjes JA, et al. Eur Urol. 2024;85(1):17-31. **5.** Dilts D, et al. Med Decis Making. 1995;15(4):333-47. **6.** McLachlan GJ. Stat Methods Med Res. 1992;1(1):27-48. **7.** Romesburg HC. *Cluster analysis for researchers*. Belmont: Lifetime Learning Publications; 1984. **8.** Niegisch G, et al. Future Oncol. 2024;20(19):1351-66. **9.** Niegisch G, et al. Ann Oncol. 2023;34(Suppl 2):Abstract 2386P.

DISCLOSURES M. Kearney is an employee of Merck and owns stock and has other ownership interests in Merck, Novartis, and UCB. **F. Hardtstock** was an employee of Cytel at the time of the study and has served in a consulting or advisory role for Merck. **J. Krieger** was an employee of Cytel at the time of the study and has served in a consulting or advisory role for Merck. **B. Deiters** is an employee of GWQ ServicePlus AG. **U. Maywald** was an employee of AOK PLUS at the time of the study. **U. Osowski** is an employee of Merck Healthcare Germany GmbH, Weiterstadt, Germany, an affiliate of Merck KGaA, and holds stock and other ownership interests in Merck. **T. Wilke** is an employee of IPAM e. V., and is a scientific lead and partner at GIPAM. **G. Niegisch** reports talks and symposia for Astellas, AstraZeneca, Bristol Myers Squibb, medac, Pfizer, and Roche; has served on advisory boards for Bristol Myers Squibb, Ipsen, Janssen, Merck, Pfizer, Roche, and Sanofi; and has received travel and accommodation expenses from AstraZeneca, Merck, Pfizer, and Roche. **M.-O. Grimm** has served in a consulting or advisory role for Astellas Pharma, AstraZeneca, Bayer/Vital, Bristol Myers Squibb, Eisai, EUSA Pharma, Ipsen, Merck, MSD, Pfizer, Roche Pharma AG, and Takeda; has received travel, accommodations, and expenses from Bristol Myers Squibb and Merck; has received honoraria from Astellas Pharma, AstraZeneca, Bristol Myers Squibb, EUSA Pharma, Ipsen, MSD, and Pfizer; and has received research funding from Bristol Myers Squibb and Intuitive Surgical.

ACKNOWLEDGMENTS The authors thank the employees of AOK PLUS and GWQ for their support in providing the anonymized datasets. This study was sponsored by Merck (CrossRef Funder ID: 10.13039/100009945). Editorial support was provided by Nucleus Global and was funded by Merck. Some parts of these analyses were previously presented as a poster presentation at ESMO 2023.[9]

Poster No. MSR157. Presented at ISPOR Europe, 17-20 November 2024; Barcelona, Spain.