

13 Using Digital Longitudinal Health Care Data to Improve Health Care Quality

Herrmann, Wolfram J. ¹

Napierala, Hendrik ¹

Henschke, Cornelia ²

Mendling, Jan ^{3,4,5}

¹ Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of General Practice and Family Medicine

² Technische Universität Berlin, Fachgebiet Management im Gesundheitswesen, Berlin Centre for Health Economics Research

³ Humboldt-Universität zu Berlin, Institut für Informatik

⁴ Vienna University of Economics and Business, Department of Information Systems and Operations Management

⁵ Weizenbaum-Institut Berlin

ORCID:

Herrmann, Wolfram J.: 0000-0002-9505-4911

Napierala, Hendrik: 0000-0002-1542-2877

Henschke, Cornelia: 0000-0001-5261-2330

Mendling, Jan: 0000-0002-7260-524X

Health Care as Longitudinal Process

Health care utilization does not consist of independent events such as doctor visits but of sequences of events. In a qualitative study on the nature of health care utilization in Norway and Germany (Herrmann 2017), we observed that consultations in general practice can often not be attributed to a single reason for encounter. They are rather integral components of a series of events, linked to consultations with the same doctor before or linked to consultations in the future which are already planned as follow-up consultations. Our findings also revealed structural disparities between Norway and Germany concerning referrals, information flow, and the involvement of different healthcare providers.

The observation that health care consultations are sequences led to a sequential model of health care utilization that can be formalized as directed graphs: Graphs consist of nodes, which are connected by edges. In directed graphs, these edges have a direction. In the case of health care utilization sequences, the nodes of these graphs describe patient-provider contact events. Each event refers to a specific point in time. The edges represent the flow of information either by visiting the same provider site again or by referral or medical report. Figure 1 presents a fictitious example of such a sequence with circles representing nodes of the graph.

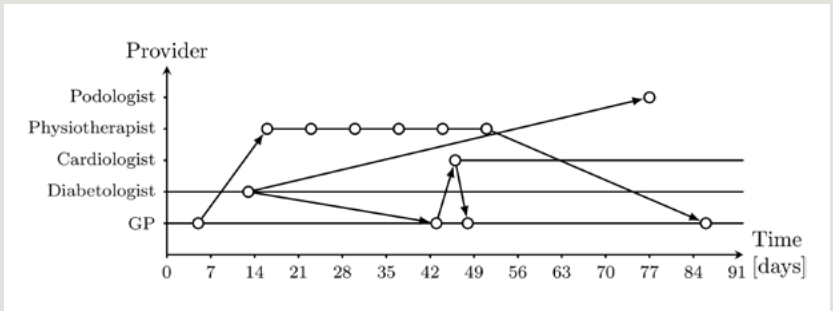


Figure 1: A fictitious example of a health care utilization sequence. (from Herrmann 2017)

Complexity of Actual Health Care Sequences in Germany

The German health care system is quite unique in offering a broad spectrum of ambulatory health care services of GPs and specialist without any formal access restrictions. Patients can choose freely which GP or specialist they want to go to. They usually are not bound

to any specific GP nor do GPs have a gatekeeping function. Hospitals are strictly separated from ambulatory health care. There are no formal patient lists and no fees to access doctors. These regulations described are valid for patients in statutory health insurance in Germany which amount for 90% of the German population.

Utilizing data from an extensive cohort of statutory German health care insurances, generously supplied by the WIDO – AOK Research Institute, we have formulated event sequences based on anonymized real patient data. We received data of 500.002 randomly chosen insurants who were insured consecutively from 2011 till 2015 at one of the regional AOKs. The data contained all diagnoses and remuneration codes of ambulatory care including the sites and health care providers with their specialty. Additionally, all hospitalizations with referral were included, all medication collected at a pharmacy and all rehabilitation if it was financed by the statutory health insurance. More information on the data can be found in Ulrich et al. 2020.

All coded contacts with patients were organized by individual doctors and practice sites; these serve as nodes. We assumed information flows inside each practice site from visit to visit, thus, each node inside a site is connected by an edge. If the code indicated a referral, we connected the according node with the most likely temporally close node which fit to the referral code. There was no data on medical reports available, thus we omitted an edge pointing back. Data on pharmacies did not include sites and could not be further distinguished. We chose to examine further sequences of patients who had a myocardial infarction in 2012-2014. Checking these sequences visually, we chose two sequences, which looked exemplary (Fig. 2) and extreme (Fig. 3).

Figure 2 shows the health care utilization pattern of a patient aged 80 to 90 years from 2011 to 2015. In 2013, the patient had a myocardial infarction. We note that this particular patient was admitted to a single hospital, albeit with considerable frequency. The patient sought care from various physicians across numerous practice locations. In 2011, there was a change in the choice of general practitioner, and subsequently, the individual visited two different general practitioners in close temporal succession.

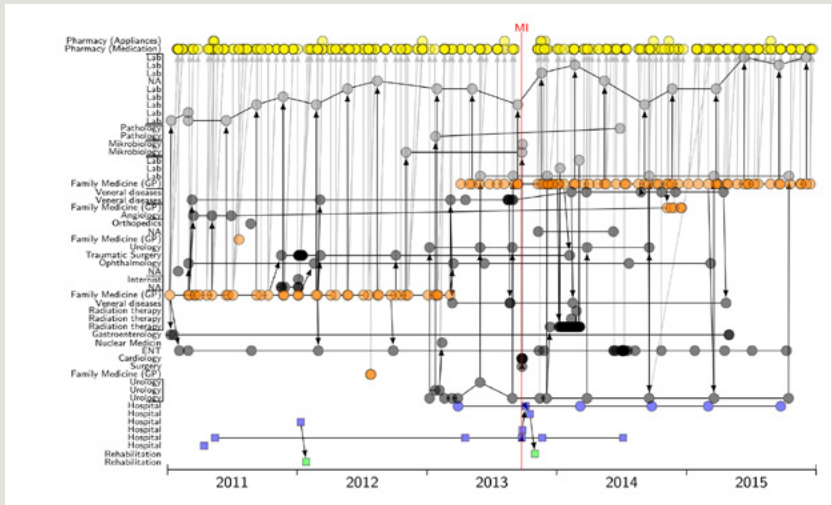


Figure 2: Real-world health care utilization sequence of a patient 80-90 years old with a myocardial infarction in 2013 in Germany

Figure 3 illustrates a complex sequence of healthcare utilization involving a patient aged between 60-70 years. This patient experienced a myocardial infarction in 2013. The depiction reveals the patient's regular changes of general practitioners and frequent visits to various doctors in the ambulatory care setting and different hospitals. This sequence vividly underscores the complexity inherent in real healthcare utilization patterns in Germany.

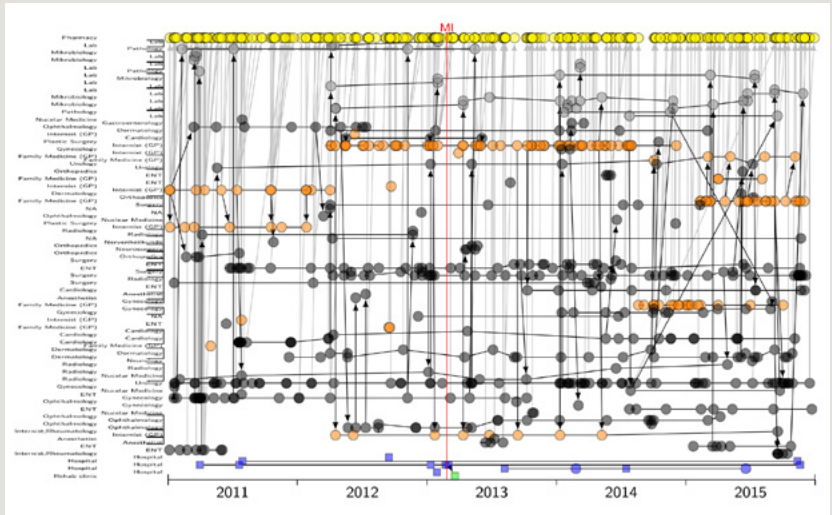


Figure 3: Real-world health care utilization sequence of a patient 60-70 years old with a myocardial infarction in 2013 in Germany

Consequences of the complexity for digitalized information systems

These exemplary health care utilization sequences show the complexity inherent in healthcare utilization patterns in Germany. These complexities carry significant implications for digitalized information systems. Given the multitude of contacts and diverse healthcare providers involved, an electronic patient record as it is currently developed in Germany that merely appends information from each consultation in plain text would prove ineffective. This sets Germany apart from healthcare systems with fewer contacts. In 2018, while the average number of face-to-face consultations with a doctor in the EU ranged between six and seven, Germany recorded around ten such consultations (OECD/European Union 2020). Consequently, there is a pressing need for alternative methods of data structuring, presentation, and visualization to afford healthcare professionals a meaningful and efficient means of accessing available healthcare data.

Visualization and analysis opportunities

There are various opportunities for visualizing, analyzing and comparing utilization sequences of individual patient data. Potential conclusions might be drawn regarding improvements for the provision of care. Several techniques for visualizing and analyzing event sequences have been defined in computer science (Yeshchenko and Mendling, 2023). For example, process mining techniques can be used to compare observed sequences with normative specifications. Corresponding techniques are known as conformance checking (van der Aalst, 2016). One challenge for the application of existing conformance checking techniques is that different utilization sequences overlap for multi-morbid patients and thus treatment paths for the individual diseases are pursued in concurrency. For that reason, standard process mining techniques are not directly applicable, but analysis techniques must be developed that can deal with overlapping sequences.

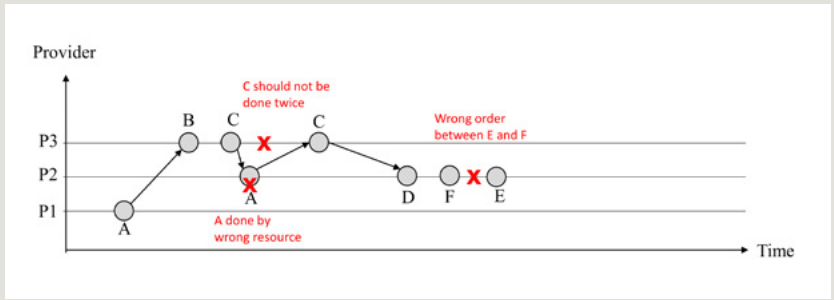


Figure 4: Behavioral rules and violations highlighted for a given health care utilization sequence

An application of conformance checking concepts could be applied to health care utilization sequences as follows. Formal rules have to be specified about which sequence of activities is expected or allowed and which providers should be the responsible resources for performing the respective activities. Deviations from this expected behavior can then be highlighted as violations in the sequence diagram.

Recent research explores the development of new analysis techniques using event graphs that define partial orders of events as an input data structure (Leemans et al., 2023). Given observed partial orders of utilization event sequences, we can compare them with recommended treatment pathways to gain valuable information for improving the quality of care. Technically, corresponding techniques have been implemented as event knowledge graphs in graph databases such as Neo4j (Esser and Fahland, 2021; Waibel et al., 2022).

Conclusion

Routine data from statutory health insurance funds are large data sources which pose unique challenges regarding their complexity (Slagman et al., 2023). Process mining techniques and various analysis techniques provide the opportunity to identify and compare utilization sequences of individual patients from routine data. Such sequence mining techniques combined with advanced visualizations can support the integration of data-driven insights into care. Their use by physicians and health insurance funds could help avoid overuse, underuse and misuse and increase the efficiency of care, e.g. health insurers are able to coordinate complex patients by providing case managers. Nevertheless, there is a need for further interdisciplinary research, particularly on the implementation of those techniques.

References

- Herrmann, WJ (2017), Haarmann A, Bærheim A (2017)** A sequential model for the structure of health care utilization. *PLOS ONE* 12(5): e0176657. <https://doi.org/10.1371/journal.pone.0176657>
- Yeshchenko, A. and Mendling, J. (2024)**. A survey of approaches for event sequence analysis and visualization. *Information Systems*, 120, pp.102283–102283. doi:<https://doi.org/10.1016/j.is.2023.102283>.
- van der Aalst, W. (2016)**. *Process Mining*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:<https://doi.org/10.1007/978-3-662-49851-4>.
- Sander, Zelst, van and Lu, X. (2022)**. Partial-order-based process mining: a survey and outlook. *Knowledge and Information Systems*, 65(1), pp.1–29. doi:<https://doi.org/10.1007/s10115-022-01777-3>.
- Waibel, P., Pfahlsberger, L., Revoredo, K. and Mendling, J. (2022)**. Causal Process Mining from Relational Databases with Domain Knowledge. *arXiv (Cornell University)*. doi:<https://doi.org/10.48550/arxiv.2202.08314>.
- Esser, S. and Fahland, D. (2021)**. Multi-Dimensional Event Data in Graph Databases. *Journal on Data Semantics*. doi:<https://doi.org/10.1007/s13740-021-00122-1>.
- OECD/European Union (2020)**. *Health at a Glance: Europe 2020: State of Health in the EU Cycle*, OECD Publishing, Paris, <https://doi.org/10.1787/82129230-en>.
- Ulrich, R., Pischon, T., Robra, B.-P., Freier, C., Heintze, C. and Herrmann, W.J. (2020)**. Health care utilisation and medication one year after myocardial infarction in Germany – a claims data analysis. *International Journal of Cardiology*, 300, pp.20–26. doi:<https://doi.org/10.1016/j.ijcard.2019.07.050>.
- Slagman, A., Hoffmann, F., Horenkamp-Sonntag, D., Swart, E., Vogt, V. and Herrmann, W.J. (2023)**. Analyse von Routinedaten in der Gesundheitsforschung: Validität, Generalisierbarkeit und Herausforderungen. *Zeitschrift für Allgemeinmedizin*, 99(2), pp.86–92. doi:<https://doi.org/10.1007/s44266-022-00004-0>.