



Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains



Aniek F. Markus^{a,*}, Katia M.C. Verhamme^{a,b}, Jan A. Kors^a, Peter R. Rijnbeek^a

^a Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands

^b Department of Bioanalysis, Ghent University, Ghent, Belgium

ARTICLE INFO

Article history:

Received 24 December 2021

Revised 4 August 2022

Accepted 21 August 2022

Keywords:

Treatment pathways

Current treatment practices

Observational data

R package

ABSTRACT

Background and objectives: There is an increasing interest to use real-world data to illustrate how patients with specific medical conditions are treated in real life. Insight in the current treatment practices helps to improve and tailor patient care, but is often held back by a lack of data interoperability and a high-level of required resources. We aimed to provide an easy tool that overcomes these barriers to support the standardized development and analysis of treatment patterns for a wide variety of medical conditions.

Methods: We formally defined the process of constructing treatment pathways and implemented this in an open-source R package TreatmentPatterns (<https://github.com/mi-erasmusmc/TreatmentPatterns>) to enable a reproducible and timely analysis of treatment patterns.

Results: The developed package supports the analysis of treatment patterns of a study population of interest. We demonstrate the functionality of the package by analyzing the treatment patterns of three common chronic diseases (type II diabetes mellitus, hypertension, and depression) in the Dutch Integrated Primary Care Information (IPCI) database.

Conclusion: TreatmentPatterns is a tool to make the analysis of treatment patterns more accessible, more standardized, and more interpretation friendly. We hope it thereby contributes to the accumulation of knowledge on real-world treatment patterns across disease domains. We encourage researchers to further adjust and add custom analysis to the R package based on their research needs.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Background

There is an increasing interest to use real-world data to illustrate how patients with specific medical conditions are treated in real life. Real-world data is already used for post-market safety surveillance of medicines, to monitor drug consumption and related costs, to detect inappropriate prescribing and to assess compliance with treatment guidelines [1]. This research focusing on treatment patterns allows to analyze drug utilization beyond the volume of drug uptake (i.e. incidence and prevalence numbers on prescribing/dispensing), as it allows to investigate treatment pathways (i.e. first-line and subsequent treatments over time). The latter is crucial to investigate adherence to treatment guidelines in their entirety. Clinical guidelines are systematically developed recommendations informed by a review of scientific evidence and an

assessment of the benefits and harms of alternative treatments for a specific medical condition [2]. These guidelines therefore support evidence-based medicine by serving as a 'best practice'. Unfortunately, clinical guidelines are often not well adhered to in practice for a variety of reasons including lack of applicability, organizational constraints, or lack of awareness [3].

Knowledge of the current treatment practices is an important starting point to improve clinical practice and from a community perspective to facilitate the rational use of drugs. Evidence from real-world data has driven changes in clinical guidelines, but it is recognized that this data is not yet used to its full potential [1]. An increasing amount of routine health care data is available in the form of electronic health records, claims data and registries, which allows the analysis of treatment patterns in real life [4–7]. Currently, there is no easy-to-use tool available that is still flexible enough to accommodate specific research needs. Barriers to analyze treatment practices might be a lack of data interoperability and a high-level of required resources (i.e. amount of time and level of expertise).

* Corresponding author.

E-mail address: a.markus@erasmusmc.nl (A.F. Markus).

In this paper we present TreatmentPatterns, an open-source R package to enable the analysis of treatment patterns that overcomes these barriers. The Observational Medical Outcomes Partnership (OMOP) common data model (CDM) maps data in different structures, formats and terminologies to one common standard to promote interoperability of health care data [8]. This allows to perform multi-database studies without the need to share patient-level data. Moreover, the OMOP CDM allows for the development of standardized tools, which is important to produce reproducible and timely real-world evidence. A wide range of tools is already available to support research on the OMOP CDM; Health Analytics Data-to-Evidence Suite (HADES)¹ is a set of open source R packages for large-scale analytics, including population characterization, population-level causal effect estimation, and patient-level prediction. It also provides supporting packages for new analytical pipelines. These building blocks are used by TreatmentPatterns to analyze treatment patterns of a study population of interest. The package can be executed against the OMOP CDM, but the main parts of the package are also usable with other data formats. To our knowledge, no earlier work formally defined and implemented the process of constructing treatment pathways from medical files.

Objectives

We aimed to provide a user-friendly tool to support the development and analysis of treatment patterns for a wide variety of medical conditions, thereby making the analysis of treatment patterns more accessible, more standardized, and more interpretation friendly. We demonstrated the package by analyzing the treatment patterns of three common chronic diseases (type II diabetes mellitus, hypertension, and depression) using data from the Dutch Integrated Primary Care Information (IPCI) database [9].

Methods

In this section, we first formally define the problem of constructing treatment pathways following earlier work [4]. Based on the problem formalization, we then describe the implementation of the open-source R package TreatmentPatterns, which enables a reproducible and timely analysis of treatment patterns.

Problem formalization

We defined the problem as follows: for a specified study population (i.e. target cohort), the goal is to find the treatment pathway consisting of selected “treatments of interest” (i.e. event cohorts). Treatments of interest (i.e. events) can be drug prescribing and/or dispensing, but could also be other therapies, procedures, or measurements. The treatment pathway is defined as the sequence of treatments of interest over time. The target cohort is the set of patients for whom we want to study treatment patterns and should be defined by a clear set of inclusion/exclusion criteria. Similarly, event cohorts specify which patients have received a given treatment of interest. We use the term event era to refer to the span of time a patient is exposed to a specific treatment of interest. How to define target and event cohorts is elaborated on in other sources (e.g. Chapter 10 ‘Defining cohorts’ in the Book Of OHDSI [10]). The challenge addressed in this paper is how the treatments of interest should be processed to construct treatment pathways.

For each individual in a target cohort, we extract the treatments from his or her medical file after the index date. Note that individuals can have many different, sometimes simultaneously occurring

treatments in their medical file. The index date is the moment in time the individual enters the target cohort. Fig. 1 shows an example medical file and visualizes the decisions that need to be made to construct the treatment pathway for an individual receiving treatments A, B and C.

To create treatment pathways, the following decisions need to be made:

1. Lookback period treatments (periodPriorToIndex)

The first decision that needs to be made is the moment in time from which selected treatments of interest should be included in the treatment pathway. The default is all treatments starting after the index date of the target cohort. For example, for a target cohort consisting of newly diagnosed patients, treatments after the moment of first diagnosis are included. However, it might also be desirable to include (some) treatments prior to the index date, for instance in case a specific disease diagnosis is only confirmed after initiating treatment. Therefore, *periodPriorToIndex* specifies the period (i.e. number of days) prior to the index date from which treatments should be included.

2. Minimum duration treatment (minEraDuration)

The duration of extracted event eras may vary a lot and it can be preferable to limit to only treatments exceeding a minimum duration. Hence, *minEraDuration* specifies the minimum time an event era should last to be included in the analysis.

3. Maximum time interval between treatments (eraCollapseSize)

If an individual receives the same treatment for a longer period of time (e.g. need of chronic treatment), one is likely to need refills. As patients are not 100% adherent, there might be a gap between two subsequent event eras. Usually, these eras are still considered as one treatment episode and the *eraCollapseSize* defines the maximum gap within which two eras of the same event cohort would be collapsed into one era (i.e. seen as continuous treatment instead of a stop and re-initiation of the same treatment).

4. Minimum overlap of different treatments to be considered as combination treatment (combinationWindow)

Patients often receive different treatments at the same time, which may occur with full overlap or partial overlap. If the period of overlap is short, this is defined as a *switch* in treatment, which means that the earlier treatment is switched to a new treatment. If the overlap is longer, we assume that both treatments are received at the same time. To differentiate between the two options, *combinationWindow* specifies the time that two event eras need to overlap to be considered a combination treatment. If there are more than two overlapping event eras, we sequentially combine treatments, starting from the first two overlapping event eras.

5. Minimum duration of post combination treatment (minPostCombinationDuration)

As a result of creating combination treatments, there might be short remaining periods of a single treatment before or after the combination treatments. For similar reasons as described above, we might only want to consider treatments if they last for a sufficient period of time. Therefore, *minPostCombinationDuration* defines the minimum time that an event era before or after a generated combination treatment should last to be included in the pathway as a separate treatment.

6. Select which treatments should be included in pathway (filterTreatments)

Depending on the research question, we might be interested to include repeated treatments in the pathway or not. The *filterTreatments* parameter allows to specify whether to include all

¹ <https://github.com/OHDSI/Hades>.

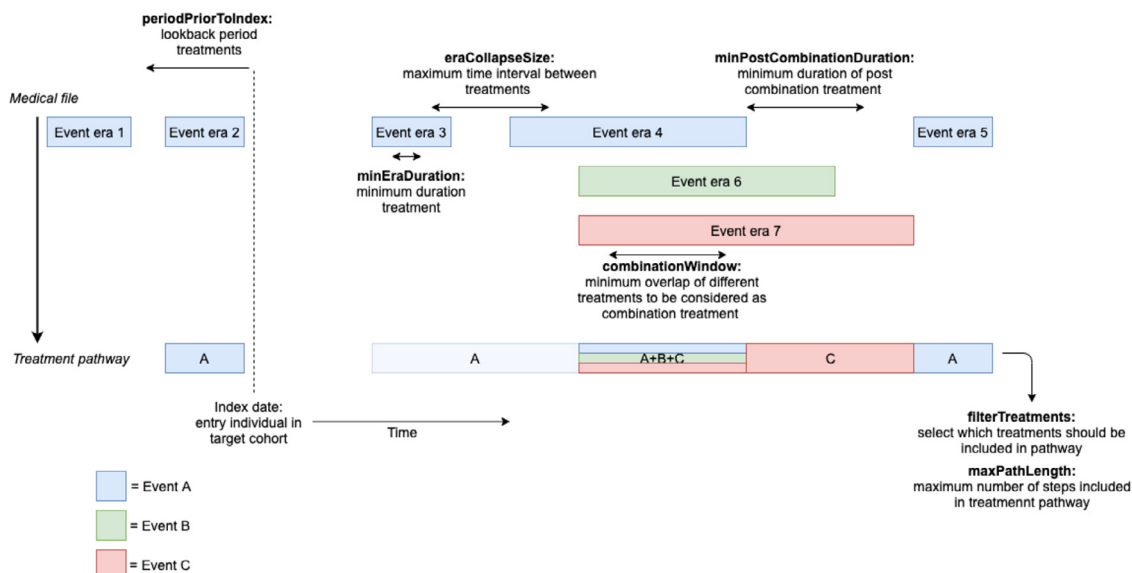


Fig. 1. Summary of decisions to construct individual treatment pathways. The events A, B, and C refer to the treatments of interest. For full detailed steps from medical file to treatment pathway see Appendix A.

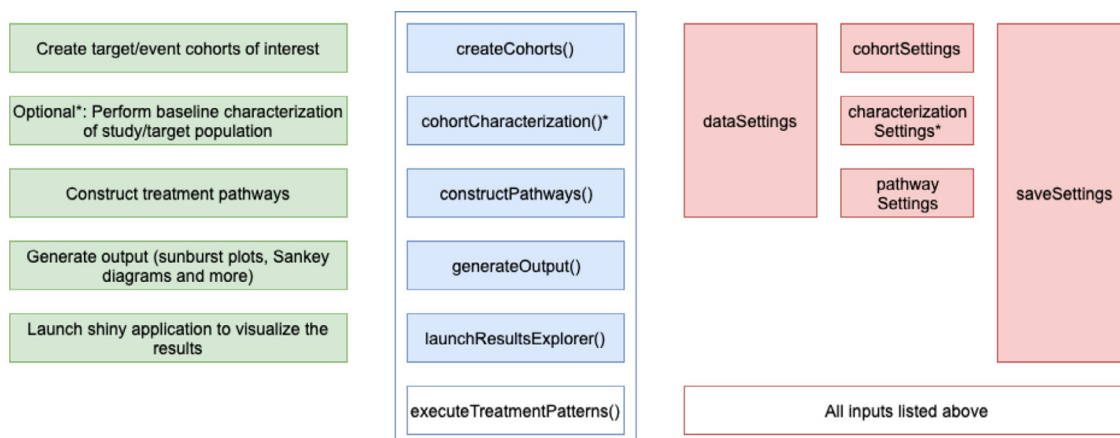


Fig. 2. Structure of R package TreatmentPatterns. * Only for data in OMOP CDM format.

treatments ('All'; A - A - A+B+C - C - A), to remove sequential repeated treatments ('Changes'; A - A+B+C - C - A), or to only include first time occurrences of treatments ('First'; A - A+B+C - C).

7. Maximum number of treatments included in pathway (maxPathLength)

Finally, maxPathLength specifies the maximum number of treatments included in the treatment pathway (i.e. the number of layers in a sunburst plot).

Implementation

The R package TreatmentPatterns implements the problem formalization described above and is structured as shown in Fig. 2. The package allows to run the full analysis at once with the executeTreatmentPatterns function or by sequentially running the individual subfunctions (i.e. createCohorts, cohortCharacterization (optional for databases mapped to the OMOP CDM), constructPathways, generateOutput, and launchResultsExplorer).

The main inputs of the package are the target and event cohorts of interest, which can be cohort definitions to allow automatic cohort extraction from databases mapped to the OMOP CDM or a csv file to directly import existing target and event cohorts

(to be specified in cohortSettings). The package also allows to perform a baseline characterization of the study population including a wide range of pre-specified covariates (e.g. age/gender, Charlson comorbidity index score, comorbid conditions with different look-back periods), which can be specified in characterizationSettings. Moreover, it is possible to add custom covariates using SQL code. Furthermore, the pathwaySettings for the construction of treatment pathways need to be specified. In addition to the settings related to the creation of the individual treatment pathways above, these include output settings to aggregate the treatment pathways (e.g. minCellCount, groupCombinations). For the full list of pathway settings with description and expected input we refer to Appendix B. Other necessary inputs are dataSettings and saveSettings, which indicate the location of the input data and output respectively.

After cohort extraction (and optional baseline characterization of the study population), the individual treatment pathway for each patient is constructed as follows:

1. Pre-process the medical file according to, amongst others, periodPriorToIndex, minEraDuration, and eraCollapseSize.
2. Identify which event eras have overlap for each person and select the first two overlapping event eras (of different events) per person.

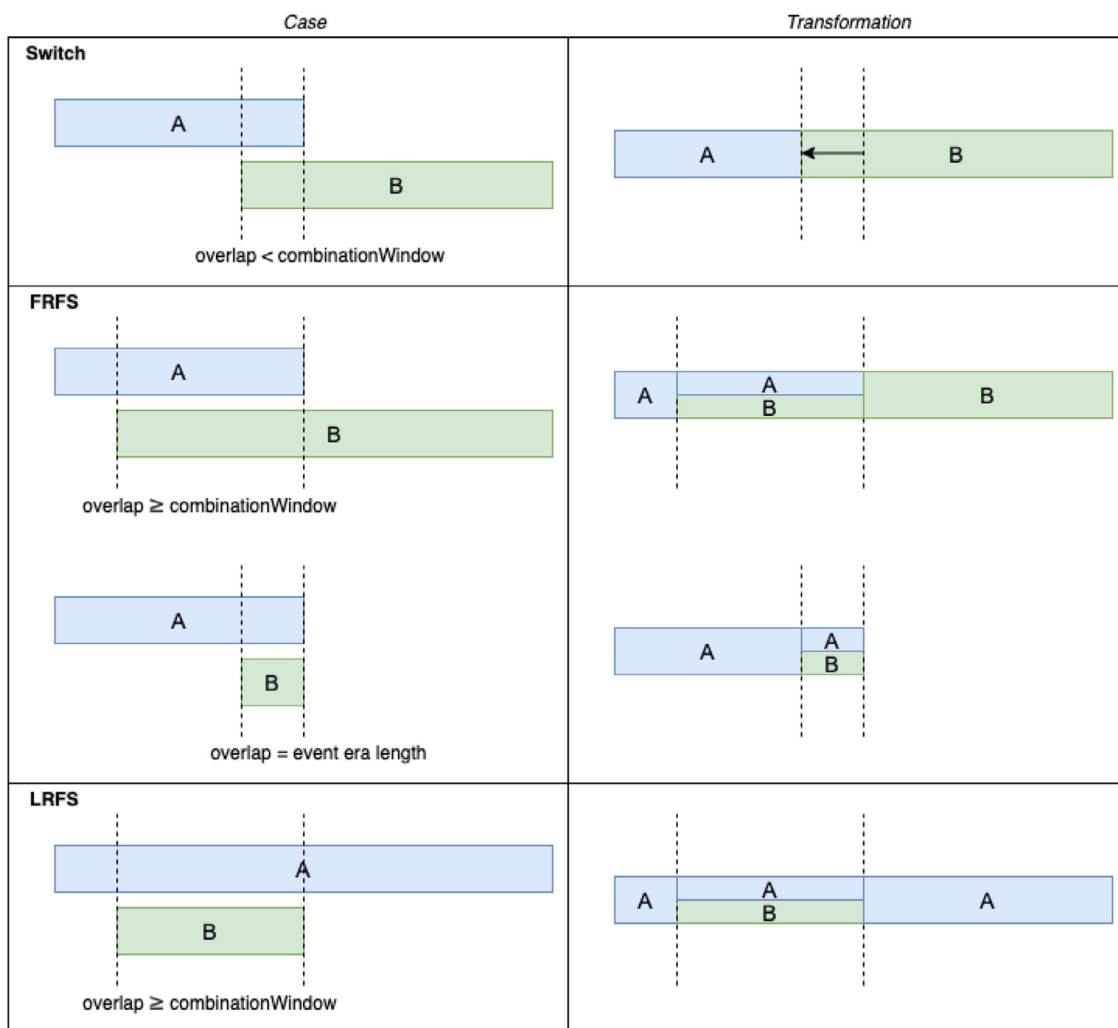


Fig. 3. Three cases of overlap can occur: a) *Switch*, b) *FRFS* ('First Received, First Stopped'), and c) *LRFS* ('Last Received, First Stopped').

3. We now distinguish three cases of overlap (see Fig. 3):
 - a. If the overlap of the two event eras is smaller than *combinationWindow* AND the overlap is not equal to the total length of one of the event eras, then this is considered a *Switch*. The end date of the earlier event will be adapted to the start of the later event.
 - If the overlap of the two event eras is at least *combinationWindow* OR the overlap is equal to the total length of one of the events, this is considered:
 - b. *FRFS* ('First Received, First Stopped') if the end date of the earlier starting event era is before or at the end date of the later starting event era.
 - c. *LRFS* ('Last Received, First Stopped') if the end date of the earlier starting event era is after the end date of the later starting event era.
4. Filter the event eras before or after the generated combination treatments based on *minPostCombinationDuration* and go back to Step 2. Stop if there are no more event eras with overlap.

After constructing the individual treatment pathways, the results are aggregated for each target cohort. The package provides several functions for displaying the results of the treatment pathways analysis in both graphical and tabular formats. The package creates sunburst plots, Sankey diagrams, and various other outputs (e.g. percentage of people treated, treatment changes over time, average duration of event eras) to give insight in first-line,

second-line, and higher-line treatments. *Sunburst plots* of treatment pathways show the first treatment in the center and subsequent treatments in the surrounding outer layers. Each color represents a treatment (as defined by the event cohorts) and a layer with multiple colors indicates a combination therapy. This shows the full treatment pathway of patients over time. *Sankey diagrams* also visualize treatment pathways, but show follow-up treatment conditional on the previously received treatment by directed arrows from left to right. The size of the arrows is proportional to the number of patients. All results can be explored in an interactive Shiny application.

For full details and instructions to use the package (including a package manual and vignette) we refer to GitHub: <https://github.com/mi-erasmusmc/TreatmentPatterns>.

Results

To demonstrate the functionality of the package, we present an example study analyzing treatment patterns of three common chronic diseases using data from the Dutch IPCI database. The IPCI database contains records of more than two million patients enrolled at selected general practitioners (GPs) throughout the Netherlands [11]. We aimed to study treatment patterns of patients with type II diabetes mellitus, hypertension, and depression.

Table 1
Overview of example study on type II diabetes mellitus, hypertension, and depression.

	Type II diabetes mellitus	Hypertension	Depression
Target cohorts*	Adults with condition occurrence of: - Diabetes mellitus (201820) AND - No type 1 diabetes (35506621)	Adults with condition occurrence of: - Hypertensive disorder (316866)	Adults with condition occurrence of: - Depressive disorder (440383) AND - No schizophrenia (435783) - No bipolar disorder (432876)
Event cohorts*	Alpha-glucosidase inhibitors Biguanides Dipeptidyl peptidase 4 inhibitors (DPP-4) Glucagon-like peptide-1 receptor agonists (GLP-1) Insulin Meglitinides Other anti-diabetics Sodium-glucose co-transporter-2 inhibitors (SGLT2) Sulfonylureas Thiazolidinediones	Angiotensin-converting enzyme inhibitors (ACE) Alpha-1 blockers Angiotensin receptor blockers (ARBs) Beta blockers Calcium channel blockers (CCBs) Direct vasodilators Diuretics	Norepinephrine-dopamine reuptake inhibitors (NDRIs) Selective serotonin reuptake inhibitors (SSRIs) Serotonin antagonist and reuptake inhibitors (SARIs) Serotonin-norepinephrine reuptake inhibitors (SNRIs) Tetracyclic antidepressants Tricyclic antidepressants (TCAs)
Baseline characteristics	Age, sex, Charlson Comorbidity Index score, and a history of the following conditions: cerebrovascular disease, depressive disorder, diabetes mellitus, heart failure, hypertensive disorder, ischemic heart disease, kidney disease, obesity		
Pathway settings	periodPriorToIndex = 0 days minEraDuration = 5 days eraCollapseSize = 30 days combinationWindow = 30 days minPostCombinationDuration = 30 days filterTreatments = Changes, remove sequential repeated treatments maxPathLength = 5 minCellCount = 5 minCellMethod = Adjust groupCombinations = 10 addNoPaths = TRUE		
Data source	Integrated Primary Care Information (IPCI), Netherlands		

* For more details and code lists we refer to Appendix C.

Step 1: define target/event cohorts

We created three target cohorts of patients with index date at first diagnosis of the respective disease. Patients needed to have at least 365 days prior observation database time and 1,095 days follow-up time after index date (to allow sufficient time to investigate treatment patterns). We only included adult patients with a first diagnosis since 2010. The event cohorts include frequently prescribed drugs for each of the chronic diseases. An overview of the study is provided in Table 1, for more details and code lists we refer to Appendix C.

Step 2: specify baseline characteristics of interest

We chose to include the following baseline characteristics of interest: age, sex, Charlson Comorbidity Index score, and a history of the following conditions: cerebrovascular disease, depressive disorder, diabetes mellitus, heart failure, hypertensive disorder, ischemic heart disease, kidney disease, and obesity. For each target cohort we excluded the history of the corresponding disease from the list of baseline characteristics.

Step 3: specify settings to construct treatment pathways

We constructed treatment pathways from the moment of first diagnosis, including the treatments of interest as specified for the different study populations in Table 1. We restricted the analysis to event eras that last at least 5 days (*minEraDuration*) and considered a treatment to be a combination if the overlap is 30 days (*combinationWindow*). We included all changes between treatments (*filterTreatments*), thereby removing sequential repeated treatments but keeping recurring treatments in the pathways (i.e. A – A+B+C – C – A). For the full list of pathway settings see Table 1.

Step 4: execute study

The complete treatment patterns study was executed by running the *executeTreatmentPatterns* function after defining the setting objects (see package vignette for a code example). This setup has the advantage that limited programming expertise is required as the package will automatically execute all steps and produce an interactive Shiny application, the results of which are subsequently discussed.

Step 5: check out results

The results of the example study are available at: <https://mi-erasmusmc.shinyapps.io/TreatmentPatterns/>. We found 50,285 type II diabetes mellitus patients, 120,675 hypertension patients, and 32,567 patients with depression. Baseline characteristics showed that the proportion of males is slightly lower than that of females in patients newly diagnosed with hypertension (44.2%) and depression (35.8%) whereas for type II diabetes mellitus no female predominance was observed. The interface allows to easily compare baseline characteristics across databases or study populations (as shown in Fig. 4).

Treatment patterns were visualized in the form of various plots and an example of a sunburst plot is shown in Fig. 5 (for remaining plots see Appendix D). For type II diabetes mellitus, biguanides were the most frequently prescribed first-line treatments (62.4% of treated patients) followed by insulin (9.9% of treated patients). GPs thus seems to align with regard to the type of first-line treatment for type II diabetes mellitus. The proportion of switching was low as 35.2% of treated patients remained on biguanides (i.e. did not receive another type of treatment afterwards). More variability in the type of second-line treatments was observed, and this also included more combination treatments (62.3% as opposed to

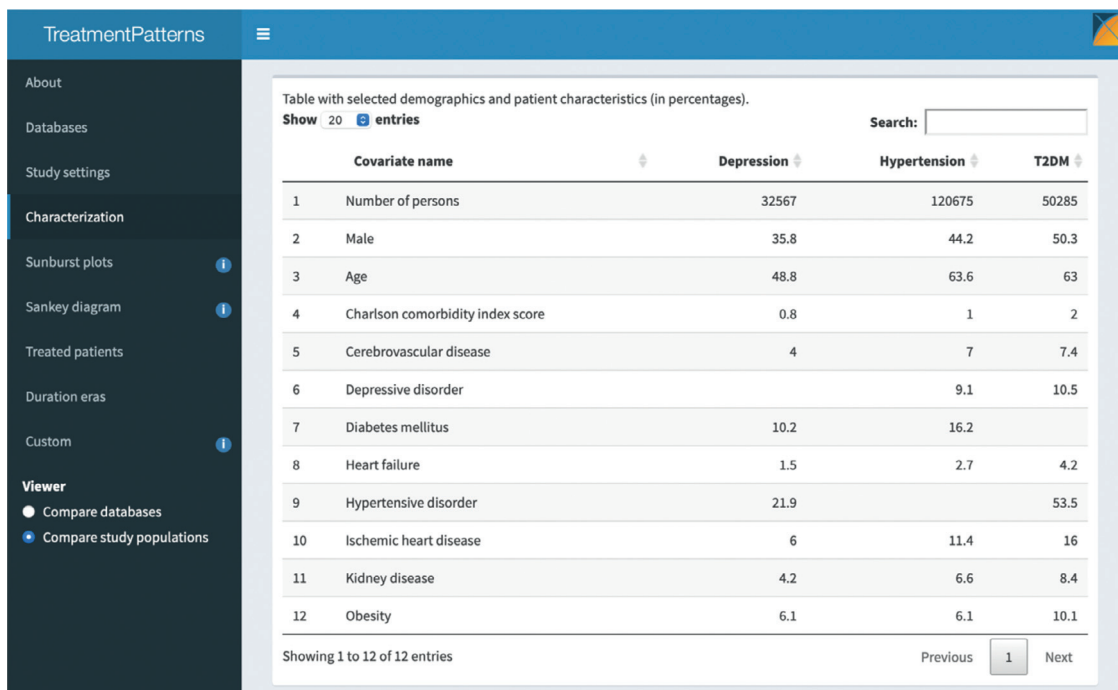


Fig. 4. Baseline characterization of study populations in IPCI.

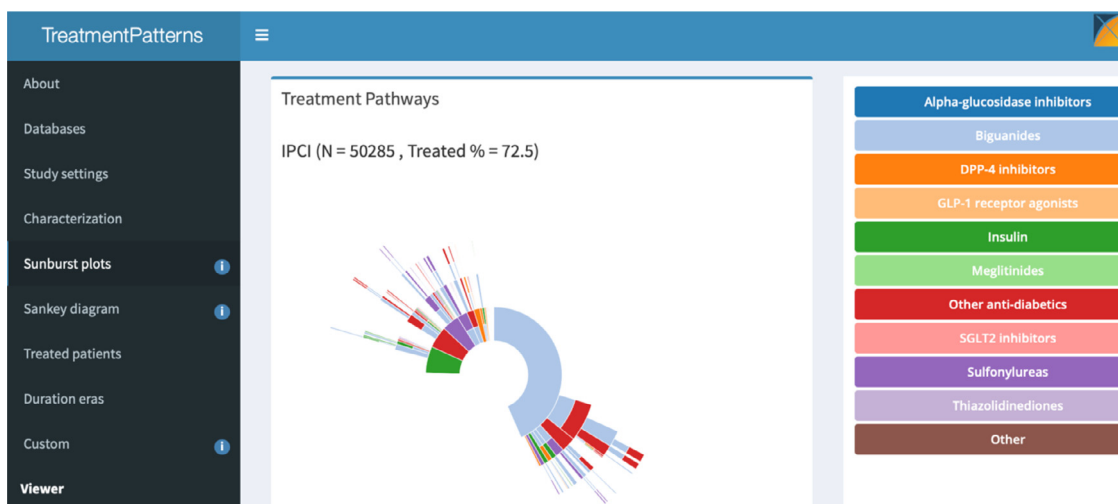


Fig. 5. Sunburst plot visualizing the treatment pathways of patients with type II diabetes mellitus in IPCI.

11.1%). For hypertension, ACE inhibitors (26.7% of treated patients), beta blockers (19.1% of treated patients), CCBs (17.1% of treated patients) and ARBs (16.5% of treated patients) were all frequently prescribed as first treatment (see Fig. D.1). For depression, SSRIs were the most frequently prescribed first-line treatment (61.2% of treated patients), followed by three types of medication that were each received by at least 10% of the treated patients: SNRIs, TCAs, and tetracyclic antidepressants (see Fig. D.2). Interestingly, patients treated for depression often did not proceed to a second-line treatment (only 24.8%). Use of combination treatments as first-line treatment was mainly observed in patients with hypertension (15.0%) and type II diabetes mellitus (11.1%), but was rare in patients with depression (1.9%).

Finally, the heatmaps indicating the duration of event eras (see Fig. D.6-D.8) showed that the mean duration of first-line treatment

for type II diabetes mellitus was 661 days, for hypertension 674 days, and for depression 658 days.

Discussion

In this paper, we formalized the construction of treatment pathways and presented an R package implementing the presented analysis. This allows researchers to give insight in treatment patterns of a specified study population (target cohort) and selected treatments of interest (event cohorts) in a reproducible and timely manner. TreatmentPatterns facilitates the execution of these studies at a large-scale, i.e. across data sources, countries, and disease domains, while allowing to easily adjust settings and perform sensitivity analysis to investigate the robustness of the results (both within and between data sources).

Insight in the current treatment practices helps to improve clinical practice. It gives insight in the choice of treatment and prefer-

ences of doctors taking into account patient's phenotype characteristics. In addition, it provides indirect information on pharmaceutical expenditure and potential health care budget constraints. Furthermore, this provides knowledge on the extent to which clinical guidelines are adhered to, which is an important step in further implementing and/or revisiting current guidelines.

Previous work analyzing treatment patterns is often addressing specific use cases and includes for example the assessment of current treatment practices of patients with type II diabetes mellitus [4–7], hypertension [4], depression [4–5], and epilepsy [6]. Some examples of methodological work on treatment patterns exist, e.g. Chen, et al. [12] use clustering of patients to extract so-called 'typical treatment patterns' based on different similarity measure methods, and Yue, et al. [13] present different visualization techniques to present regimen sequences. However, neither of these studies describe how medical files - often containing information on different treatments of which some are prescribed simultaneously - can be combined into treatment pathways to summarize the diverse range of treatment patterns observed within a study population of interest. The already existing 'Cohort Pathways' tool², which was used for some of the before-mentioned studies, has the disadvantage that it has a very limited number of settings and is not customizable to address specific research needs. The presented R package is an addition to the existing base of available methods as it offers flexibility by an increased number of pathways settings and the possibility to add custom analysis.

TreatmentPatterns is a user-friendly tool to support the development and analysis of treatment patterns without requiring much programming. Therefore, the target audience is large and includes - but is not limited to - researchers, clinicians, the pharmaceutical industry, regulatory bodies, and health technology assessment agencies. The quality of the analysis is dependent on the target and event cohort definitions (i.e. phenotypes), which need to be created and validated by the user (e.g. using the OHDSI CohortDiagnostics tool). However, this is a limitation related to the quality of the data being used that holds for all types of observational data studies and is not specific for research investigating treatment patterns. Finally, it should be noted that the analysis of treatment patterns in observational data is limited to prescribing and/or dispensing data. Hence, it is not possible to infer actual treatment intake. This should be considered when interpreting the results.

In conclusion, the example study demonstrates that this tool makes the analysis of treatment patterns more accessible, more standardized, and more interpretation friendly. We hope it thereby contributes to the accumulation of knowledge on real-world treatment patterns across disease domains. Note that there might be disease-specific features of interest which are not included in the current package. We thus encourage researchers to further adjust and add custom analysis to the package based on their research needs.

Funding

This project has received support from the European Health Data and Evidence Network (EHDEN) project. EHDEN received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Conflicts of interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Observational Health Data Sciences and Informatics (OHDSI) community for building the foundation necessary to perform large-scale, international, multi-database studies through the development of the OMOP CDM and open-source tools. In particular, we like to thank the authors of the earlier paper 'Characterizing treatment pathways at scale using the OHDSI network' [4], which has provided the initial foundation and rationale to develop the TreatmentPatterns package.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2022.107081.

References

- [1] OECD. Using Routinely Collected Data to Inform Pharmaceutical Policies. 2019. <https://www.oecd.org/health/health-systems/Using-Routinely-Collected-Data-to-Inform-Pharmaceutical-Policies-Analytical-Report-2019.pdf>.
- [2] American Academy of Family Physicians. Clinical Practice Guideline Manual. 2017. <https://www.aafp.org/family-physician/patient-care/clinical-recommendations/cpg-manual.html#xiii>.
- [3] M Lugtenberg, JM Zegers-van Schaick, GP Westert, JS Burgers, Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners, *Implement. Sci.* 4 (1) (2009) 54, doi:10.1186/1748-5908-4-54.
- [4] G Hripcsak, PB Ryan, JD Duke, et al., Characterizing treatment pathways at scale using the OHDSI network, *Proc. Natl. Acad. Sci.* 113 (27) (2016) 7329–7336, doi:10.1073/pnas.1510502113.
- [5] DM Kern, MS Cepeda, F Defalco, M. Etropolski, Treatment patterns and sequences of pharmacotherapy for patients diagnosed with depression in the United States: 2014 through 2019, *BMC Psychiatry* 20 (1) (2020) 4, doi:10.1186/s12888-019-2418-7.
- [6] H Kim, S Yoo, Y Jeon, et al., Characterization of Anti-seizure Medication Treatment Pathways in Pediatric Epilepsy Using the Electronic Health Record-Based Common Data Model, *Front. Neurol.* 11 (409) (2020), doi:10.3389/fneur.2020.00409.
- [7] KA Lee, HY Jin, YJ Kim, Y-J Im, E-Y Kim, TS. Park, Treatment Patterns of Type 2 Diabetes Assessed Using a Common Data Model Based on Electronic Health Records of 2000–2019, *J. Korean Med. Sci.* 36 (36) (2021) e230–e30.
- [8] JM Overhage, PB Ryan, CG Reich, AG Hartzema, PE. Stang, Validation of a common data model for active safety surveillance research, *J. Am. Med. Inform. Assoc.* 19 (1) (2012) 54–60, doi:10.1136/amiajnl-2011-000376.
- [9] MAJ de Ridder, M de Wilde, C de Ben, et al., Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands, *Int. J. Epidemiol.* (2022), doi:10.1093/ije/dyaa026.
- [10] Observational Health Data Sciences and Informatics. The Book of OHDSI. 2019. <https://ohdsi.github.io/TheBookOfOhdsi/>.
- [11] A Vlug, J Van der Lei, BT Mosseveld, et al., Postmarketing surveillance based on electronic patient records: the IPCI project, *Methods Inf. Med.* 38 (04/05) (1999) 339–344, doi:10.1055/s-0038-1634402.
- [12] J Chen, L Sun, C Guo, Y. Xie, A fusion framework to extract typical treatment patterns from electronic medical records, *Artif. Intell. Med.* 103 (2020) 101782, doi:10.1016/j.artmed.2019.101782.
- [13] Yue L, He S, Cao J, et al. Analytical visual methods to describe practice patterns in a newly diagnosed multiple myeloma non-interventional disease registry, 2020.

² Available in ATLAS: <https://atlas-demo.ohdsi.org>.