Journal of Biomedical Informatics 148 (2023) 104546

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

A new computational workflow to guide personalized drug therapy

Simone Pernice [a,b,1], Alessandro Maglione [c,1], Dora Tortarolo [a], Roberta Sirovich [d], Marinella Clerico [c], Simona Rolla [c,*], Marco Beccuti [a,b,2], Francesca Cordero [a,b,2]

[a] Department of Computer Science, University of Turin, Corso Svizzera 185, Turin, 10149, Italy
[b] CINI Infolife laboratory, Turin, Italy
[c] Department of Clinical and Biological Sciences, Regione Gonzole 10, Orbassano, 10043, Italy
[d] Department of Mathematics, University of Turin, Via Carlo Alberto 10, Turin, 10123, Italy

ARTICLE INFO

Dataset link: https://qbioturin.github.io/connector, https://qbioturin.github.io/epimod, https://github.com/qBioTurin/GreatNector-MS

MSC:
0000
1111

Keywords:
Longitudinal data
Computational models
Multiple Sclerosis

ABSTRACT

Objective: Computational models are at the forefront of the pursuit of personalized medicine thanks to their descriptive and predictive abilities. In the presence of complex and heterogeneous data, patient stratification is a prerequisite for effective precision medicine, since disease development is often driven by individual variability and unpredictable environmental events. Herein, we present *GreatNector* workflow as a valuable tool for (i) the analysis and clustering of patient-derived longitudinal data, and (ii) the simulation of the resulting model of patient-specific disease dynamics.
Methods: *GreatNector* is designed by combining an analytic strategy composed of CONNECTOR, a data-driven framework for the inspection of longitudinal data, and an unsupervised methodology to stratify the subjects with GreatMod, a quantitative modeling framework based on the Petri Net formalism and its generalizations.
Results: To illustrate *GreatNector* capabilities, we exploited longitudinal data of four immune cell populations collected from Multiple Sclerosis patients. Our main results report that the T-cell dynamics after alemtuzumab treatment separate non-responders versus responders patients, and the patients in the non-responders group are characterized by an increase of the Th17 concentration around 36 months.
Conclusion: *GreatNector* analysis was able to stratify individual patients into three model meta-patients whose dynamics suggested insight into patient-tailored interventions.

1. Introduction

Thanks to their ability to handle complex and heterogeneous data, computational models can be instrumental in analyzing molecular processes, testing hypotheses, and obtaining a more profound understanding of the mechanisms that drive diseases. For these reasons, several computational modeling approaches have surged as promising analytical tools in pre-clinical and clinical research to investigate complex diseases in great detail with the potential of reaching personalized treatment strategies [1].

In the biological and medical fields, the analysis of collections of longitudinal data from different patients using data-driven models allows their stratification into specific groups based on the evolution of relevant biomarkers and their associated clinical surveillance. Patient stratification is a prerequisite for effective precision medicine, since disease development is often driven by individual variability and unpredictable environmental events, and, among a plethora of available therapies, the drug choice still relies on the clinician's judgment and experience based on the patients' clinical status, coexisting comorbidities and administration method preferences. Yet, a deeper understanding of how these factors concur to cause disease onset and progression requires a functional understanding of the underlying mechanisms of disease, which is better captured by mechanism-based models.

Several high-level formalisms are available for mechanism-based models analysis. Among them, the Petri Net (PN) formalism [2] is especially suitable for the creation and analysis of mechanism based-models in the biomedical field, and it has been exploited to model biological phenomena from the molecular level [3,4], to cellular [5,6] and epidemiological levels [7,8]. Thanks to their graphical notation, Petri Nets allow a compact description of the system of interest from which the underlying stochastic and deterministic processes can be automatically derived using software such as epimod [8]. Applied in personalized medicine, these mechanism-based models can be informed and parameterized by comprehensive analyses of clinically

* Corresponding author.
E-mail address: simona.rolla@unito.it (S. Rolla).
[1] Contributed equally to this work.
[2] Jointly supervised this work.

https://doi.org/10.1016/j.jbi.2023.104546
Received 28 February 2023; Received in revised form 9 November 2023; Accepted 13 November 2023
Available online 19 November 2023

derived longitudinal data to simulate patient disease phenotypes and to identify subgroups of drug responders and non-responders. Integration of patient-specific data in Boolean models has successfully simulated heterogeneity in small cell lung cancer [9] and pancreatic cancer [10], and tumor type-specific signatures of receptor proteins associated with epithelial–mesenchymal transition in the bladder and breast cancer [11], suggesting possible therapeutic targets.

Weatherley et al. [12] provides an overview of the ongoing mathematical research dedicated to enhancing our understanding of Multiple Sclerosis disease. It specifically introduces various modeling approaches to examine the disease from diverse perspectives. Additionally, in the study conducted by Maleki et al. [13], patient-specific data were integrated into an agent-based model called Universal Immune System Simulator to simulate the human immune system dynamics under multiple sclerosis. Indeed, this model was effectively employed to retrospectively validate the effects of various treatments at both the individual and population levels using data from two clinical trials. However, only a limited number of these studies take advantage of widely accessible and user-friendly computational frameworks to ensure that these approaches are accessible to physicians and researchers who may not possess advanced mathematical skills. We believe it is essential to develop a computational environment that integrates longitudinal data analysis and quantitative modeling analysis, providing and supporting users with simple and user-friendly interfaces. Here, we propose the workflow GreatNector, which combines CONNECTOR[3] [14], a data-driven framework that analyzes and inspects longitudinal data based on statistical methods for Functional Data Analysis (FDA) [15,16], with GreatMod[4] [8], a quantitative modeling framework based on the Petri Net formalism and its generalizations. In previous works, we applied GreatMod to model Multiple Sclerosis (MS), with the aim of identifying the key parameters involved in the modulation of therapy efficacy, and of performing *in silico* experiments to improve our understanding of this complex disease. In detail, we showed how GreatMod can be applied to model several scenarios considering pregnancy [17], different drug treatments [6,17], and space dependency [18]. Here, we use CONNECTOR to expand on our previous models to include individual patients' characteristics, a feature that is especially important in a complex disease such as MS, but that can be exploited in any clinical scenario. In our PN MS model, the integration of patient-derived data into a mechanistic model can provide an invaluable tool for the definition of (i) a patient-specific personalized drug administration schedule, and (ii) for the prediction of therapy efficacy based on the patient's peculiar immunological structure at the time of the onset of the disease or therapy.

2. Material and methods

A description of MS disease, the available longitudinal data, and the two computational methods used to analyze the data and predict the model outcome are reported in the following.

2.1. MS background

Genetics and environmental exposures are connected to disease risk and can impact the success of an intervention. A large inter-individual variation in disease processes exists and each individual may require tailored interventions based on their unique characteristics. Personalized interventions require biomarkers, useful for detecting changes in health status, and detailed understanding of pathological processes [19]. The identification of features that distinguish responding patients from patients who will develop relapses is one of the major challenges for physicians when setting up the best therapy for their patients. In the MS context, immune reconstitution therapies, such as alemtuzumab, induce quantitative and qualitative alterations of immune components, providing a sort of re-balancing of immune-tolerance networks. Alemtuzumab is a monoclonal antibody that binds to CD52, a protein present on the surface of mature lymphocytes, and induces a robust depletion of immune system cells followed by an immune reconstitution [20]. Data from clinical trial follow-ups show that alemtuzumab can have a durable effect in around one-half or less of the patients [21,22]. The remaining half of patients discontinued alemtuzumab treatment due to the inefficacy of therapy or to the manifestation of adverse events, including the development of secondary autoimmunity in a proportion of patients or opportunistic infections, due to the deep and long-lasting immune suppression. The reasons behind the different responses to treatment can be researched in the inter-individual variability of MS [23].

2.2. MS cohort description

For the analysis of this work, we retrieved data from a 6-year follow-up of MS patients treated with alemtuzumab [20]. The dataset is composed of 29 subjects (17 females, 12 males) with an average age of 34 ± 8.7 years. MS patients were diagnosed 5.0 ± 3.4 years before starting alemtuzumab and almost all the subjects were previously treated with $IFN-\beta$. Patients were subjected to periodical neurological assessments to collect clinical data, including the manifestations of disease progression (relapses) and disability assessment using the Expanded Disability Status Scale (EDSS) score. 13 out of 29 patients had one or more (1.6 ± 0.9) relapse(s) during the 6-year follow-up [20] with a median EDSS of 2.48 ± 1.06 at baseline and a median EDSS of 2.46 ± 1.39 at 6-year observation. EDSS score is commonly used as an outcome measure in MS studies, although it has some documented weaknesses in reliability and sensitivity to change [24]. Alemtuzumab treatment consists of two drug infusions one year apart, that induce long-term remission of MS also in the following treatment-free period. Patients can be subjected to additional infusions in case of reactivation of symptoms [25].

We analyzed multiple temporal measurements of T cell subtypes from 21 patients that received the standard two courses of alemtuzumab (5 administrations of alemtuzumab 12 mg/day at month 0 and 3 administrations at month 12). From the initial cohort of 29 patients, 3 patients were excluded for an earlier dropout and 5 patients were excluded for receiving a third dose of the drug.

The study design consisted of venous blood withdrawals at baseline (before the first alemtuzumab course), 6, 12 (before the second alemtuzumab course), 18, 24, 36, 48, 60, and 72 months. Although the goal was to obtain data for each subject at each time point, there is some missing data (i.e. missing sample). The number of collected samples for each subject ranged between a minimum of 3 and the maximum of all the nine time points (median 8; interquartile range 6.25–9) Samples obtained were used for immunological phenotyping, obtaining the quantification of $CD4^+$ T cells, Th1, Th17, and Treg cells. In this dataset, Rolla et al. [20] observed that the immune reconstitution of the $CD4^+$ subsets after therapy was characterized by a restoration of the Treg suppressor function coupled with a decrease in self-reactive myelin basic protein-specific Th17 and Th1 cells suggesting a shift toward immune tolerance and a reduction of T cell recruitment to the central nervous system [20]. Moreover, the results of this work showed that the Th17/Treg ratio was higher at baseline in non-responder patients. However, no analysis had been performed on T-cell dynamics, that could definitively contain other informative features.

2.3. CONNECTOR

CONNECTOR [14] is an R package for the unsupervised analysis of longitudinal data, i.e., time series constituted by measurements collected sequentially over time.

[3] Freely available at https://qbioturin.github.io/connector
[4] Freely available at https://qbioturin.github.io/epimod

CONNECTOR is built on the model-based approach for clustering functional data [26], which is particularly effective when observations are sparse and irregularly spaced. The CONNECTOR pipeline is comprised of three steps: (i) the *pre-processing step* consisting of the visualization and inspection of the data, (ii) the *model selection step*, in which several measures are computed properly set the free parameters of the model (i.e., the dimension of the spline basis vector and the number of clusters), and (iii) the *cluster visualization step*, where the longitudinal data are sub-plotted into CONNECTOR clusters and each time point's discriminant power is shown.

The selection of the free parameters is a crucial element of the analysis. In particular, CONNECTOR proposes to choose the dimension of the spline basis as the one corresponding to the largest cross-validated likelihood, as proposed in [27], by using a ten-fold cross-validation. Two plots aid in the choice of the number of clusters. The first is the *elbow* plot that reports the total tightness (a dispersion measure of clustering) calculated by varying the number of clusters. The second plot reports a measure of the cluster separation by extending the well-known Davies–Bouldin (DB) index to the functional setting, called the functional DB (fDB) index [14]. By inspecting the two plots, the proper number of clusters can be inferred by considering both the change of slope, called *elbow* in the first plot, and the minimization of the fDB index. Furthermore, CONNECTOR returns a consensus matrix associated with a *stability score*, which informs the user of the stability of the clusters.

All mathematical details are reported in Appendix A and in [14].

2.4. GreatMod

GreatMod is a general quantitative modeling framework for the analysis of complex biological systems. GreatMod is based on a high-level graphical formalism, the Petri Net formalism and its generalizations, allowing the modelers to simplify model design and provide an intuitive description of system behavior.

GreatMod follows the guidelines provided by the Reproducible Bioinformatics Project (http://reproducible-bioinformatics.org), a non-profit and open-source project whose aim is to provide scientists with an easy-to-use and flexible environment for the development of a reproducible workflow of analysis. Compatible with these ambitions, the GreatMod framework is composed of (i) a graphical editor, GreatSPN [28], for the modeling and analysis of complex systems using the PN formalism, (ii) a user-friendly R package named *epimod* [8], for the definition of a custom workflow of analysis, and (iii) a virtualized computational environment.

GreatSPN provides a Graphical User Interface (GUI), written in Java and thus portable on different systems and architectures, which allows users to draw models using a graphical formalism. *Epimod* provides few but essential micro-services characterizing the model analysis: from the derivation of the mathematical processes (both deterministic and stochastic) underlying the PN model generated by GreatSPN, to parameter sensitivity and calibration, and finally to system behavior simulation. Furthermore, *epimod* uses the Docker virtualization service to provide a fully operational and platform-independent environment to execute the analyses, simplifying the distribution, maintenance, and utilization of *epimod*'s functionalities.

3. Results

3.1. The GreatNector workflow

The ultimate goal of personalized medicine is the parameterization of a complex model, starting from temporal measurements of multiple variables collected from a single patient, to predict the evolution and outcome of the disease. However, at the present time, the availability of multiple longitudinal data is limited. Thus, we implemented *GreatNector* workflow to identify groups of patients whose measured variables dynamics, at the systematic level, are similar. Since in each set of patients, the measured evolution of the disease is comparable, each group is associated with a *meta-patient*, that is a model patient whose measured variables are supposed to evolve in time as the averages of the curves forming the cluster. *GreatNector* is defined by joining CONNECTOR and GreatMod as shown in Fig. 1.

Starting from longitudinal data Fig. 1A, we first explore the multiple temporal variables independently through the CONNECTOR software. For each biological measurement, CONNECTOR returns the clusters that gather patients with similar behavior together. The results are summarized, for each patient, in a vector reporting the CONNECTOR cluster label for each biological measurement. Hence, each patient is now described by a sequence which encodes the CONNECTOR outputs. Finally, the vectors of CONNECTOR clusters associated with the patients are fed to a k-means clustering algorithm. The resulting clusters are the groups that define the *meta-patients*: for each measured quantity, the center of the cluster, that is the mean curve obtained from the measurements belonging to the cluster, is considered as the curve of the *meta-patient* Fig. 1B.

The *meta-patient* measurements are then imported in GreatMod to calibrate the parameters of the computational model describing the systemic level of interest (Fig. 1C). Firstly, the predictions obtained from the model are used to verify the structure of the model, taking into consideration the well-known dynamics that must be satisfied. Then, a *what if* is implemented to obtain the temporal dynamics of all biological entities of interest under a specific condition(s). The evaluation of the dynamics by an expert can be useful to define a personalized therapy for each meta-patient (Fig. 1D). Finally, an assessment of the residual disease could be considered in order to verify disease progression and, if needed, take prompt action.

Table 1
The optimal set of parameters for each immunological cell population. p: dimension of the spline basis, G: number of clusters.

	p	G	Stability score
CD4	5	3	0.87
Th1	4	3	1
Th17	5	3	0.96
Treg	5	3	0.77

3.2. Definition of MS meta-patients

To define metapatients, we collected multiple temporal measurements of cell populations from a cohort of 21 MS patients that received two injections according to the standard regimen protocol of alemtuzumab administration. Four immunological cell populations were quantified at different time points over a span of six years, see Figure B.4. When observing the plots, there is not a noticeable difference in the dynamics of relapsed versus not relapsed MS patients. We ran CONNECTOR on each set of cell population dynamics independently. All the details regarding these analyses are reported in the Appendix B. In particular, Figures B.5, B.6, B.7, and B.8 summarize the CONNECTOR analyses for the $CD4^+$, Th17, Th1, and Treg cell populations, respectively. Table 1 reports the optimal CONNECTOR parameters (i.e., p as the dimension of the spline basis, and G as the number of clusters) for each immunological population and the stability score.

The CONNECTOR results are appreciable in Figures B.9 and B.10, where for each cell population (i.e. rows), the three clusters are colored by patient identifiers in the first figure, and by the presence (red) or not (blue) of relapse events in the latter. It is interesting to note that even though the dynamics reported in each cluster are homogeneous, as demonstrated by the stability matrices, a cluster associated with the relapse event or other anamnestic features (data not shown) is not delineated.

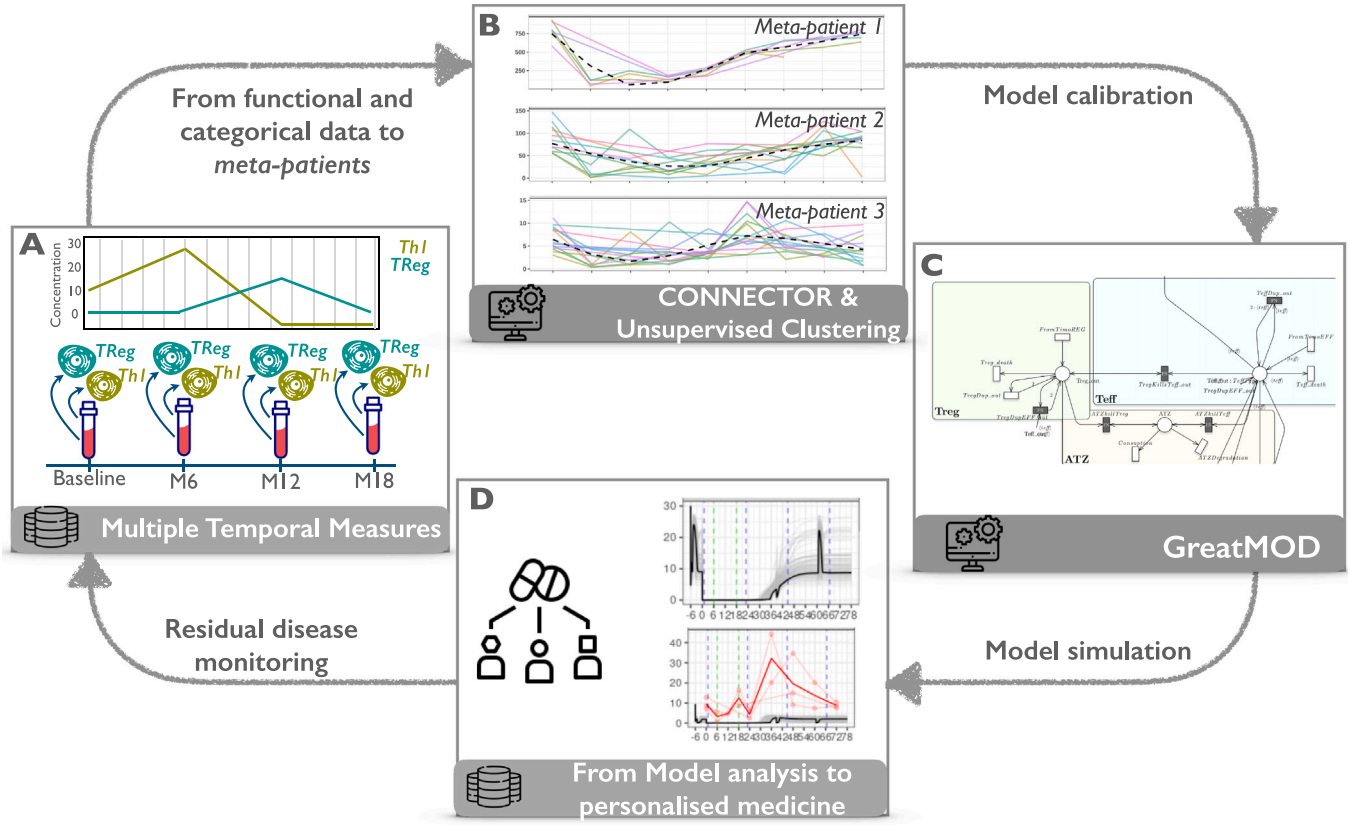

Fig. 1. GreatNector workflow. For each patient, multiple biological measurements are collected over time (A) and fed into CONNECTOR software to be stratified into meta-patients characterized by similar behavior (B). *Meta-patient* measurements are then imported in GreatMod to calibrate the parameters of the computational model (C). Model simulation and analysis are evaluated for the definition of a personalized therapy for each meta-patient (D). Finally, residual disease is monitored to assess disease progression during clinical follow-up.

Then, we ran a second clustering algorithm to identify homogeneous patient groups based on the dynamics of all the cell populations. Specifically, each MS patient is described as a list of four elements: the first is the label of the CONNECTOR cluster for the $CD4^+$, the second is the label for the Th1, the third for the Th17, and the fourth for the Treg populations (see Table 2).

Then, for each patient, the *patient-cell profile vector* is defined (e.g., from Table 2, the $CA0001$ profile is $[C, C, B, B]$). The patient-cell profile vectors are fed in a k-means clustering procedure [29] that returns three clusters. Specifically, the elbow method which measures the variability of the profile vectors within each cluster, was used to identify the optimal number of k-means clusters, reported in Appendix B. The three k-means clusters reveal three groups of patients characterized by similar cell population dynamics. We refer to these clusters as MS *meta-patients* (MSMP).

Fig. 2A shows the four immunological cell populations (rows) for each cluster (columns). The three MSMP cell population dynamics are the dashed lines. All MSMP clusters contain patients with at least one relapse event, with the exception of the MSMP_2 cluster which is characterized by the MS patients displaying two relapse events. The different trends of cell population kinetics dissected by CONNECTOR and k-means fairly predicted patients' outcomes: median TTP was not reached for MSMP_1 and MSMP_3 while median TTP was reached at 12 months for MSMP_2 (p=0.01) (Fig. 2B).

Table 2
Definition of the MS Meta-patient. The first column describes the patient ID, from the second to the fifth columns describe the *patient-cell profile vectors*, and the last column shows the K-mean cluster.

ID	$CD4^+$	Th1	Th17	Treg	K-means Cluster
CA0001	C	C	B	B	3
CA0002	C	B	B	B	3
GA0001	A	A	A	A	1
PR0001	C	A	B	B	3
PR0003	C	B	A	B	3
ZG0001	A	A	A	A	1
ZG0002	C	B	A	A	3
ZG0003	B	B	C	A	2
ZG0004	B	B	A	A	1
ZG0006	A	B	A	B	1
ZG0007	C	B	A	B	3
ZG0008	B	B	A	A	1
ZG0009	C	B	A	B	3
ZG0011	C	A	C	A	2
ZG0012	C	A	A	B	3
ZG0013	A	A	A	B	1
ZG0015	A	A	A	A	1
ZG0016	A	A	A	A	1
ZG0018	C	B	B	C	3
ZG0019	B	B	A	A	1
ZG0020	C	C	C	B	2

3.3. Model description

The PN model proposed in this work is based on the T-cell dynamics that characterize the immunopathology of relapsing remitting multiple sclerosis (RRMS). Briefly, a PN is a bipartite directed graph formed by two disjoint types of nodes: places, graphically depicted as circles, and transitions, as boxes. Places represent the state variables of the system, whose numeric value is modeled by tokens, drawn as black dots. Transitions describe the events that can induce a state change. Places and transitions are connected by arcs with a specific cardinality, which describes the number of tokens removed from or added to the

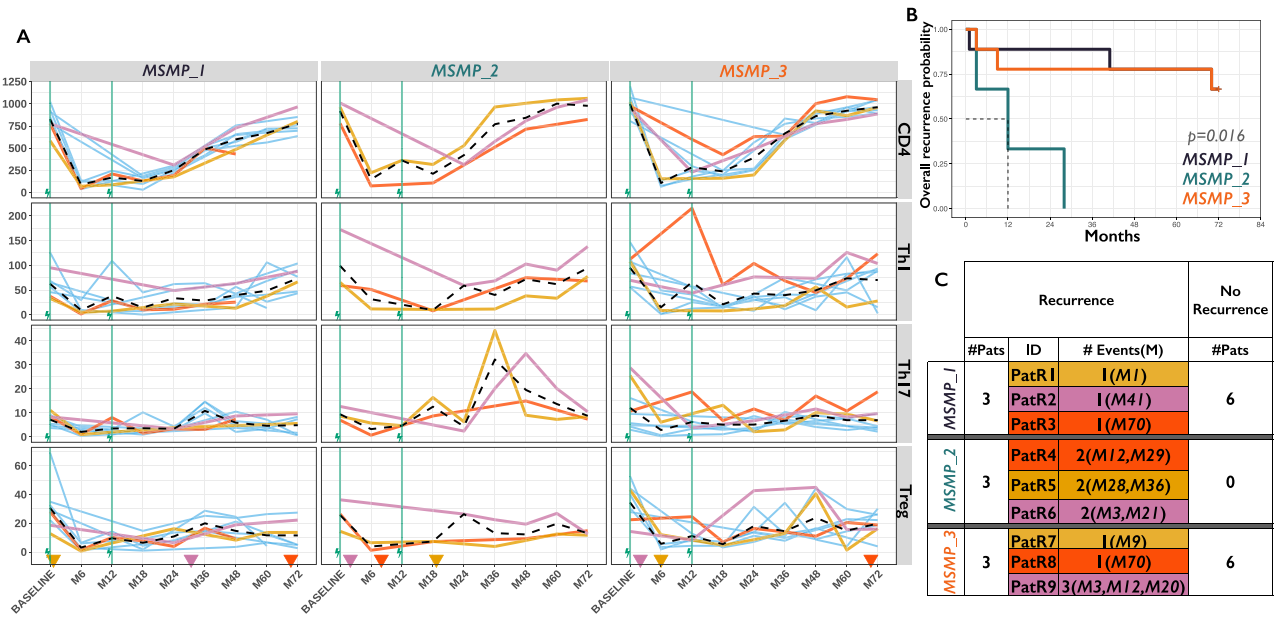

Fig. 2. (A) The columns report the three MS meta-patients (MSMPs). The rows show the four cellular population temporal dynamics. The blue lines represent the MS patients without relapse events, while the colors identify the patients characterized by at least one relapse event. The triangle symbolizes the time at which the first relapse event occurs. The dashed black lines stand for the mean curves of the corresponding MSMP and cellular population. The green vertical lines correspond to the two infusions of alemtuzumab treatment at the beginning of the therapy and at the 12th month, respectively. (B) Kaplan–Meier plot where (x-axis) represents time in months, and the vertical axis (y-axis) shows the probability of relapse occurrence. Three curves associated with the three MSMP groups highlight the different behavior of the MSMP_2 with respect to the other groups. (C) Table reporting the number of patients for each MSMP cluster (row), highlighting the number of patients per cluster without relapses (last column $\#Pats$), the number of patients with relapses, and the months of the relapse events (columns $\#Events(M)$ and the first $\#Pats$). Finally, the colors associated with each row are the line colors used in (A). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

corresponding place when a transition occurs. A transition is enabled if the number of tokens in its input places is greater than or equal to the cardinality on the corresponding arcs connecting the place to the transition. The formal definition of the PN formalism exploited in this work is reported in Appendix C. We modified the model described in Pernice et al. [6] to focus our attention on representing the dynamics of the peripheral immune system over a longer period, i.e. the 6-year follow-up in MS patients receiving alemtuzumab treatment. The model, depicted in Figure C.12 in Appendix C, consists of 7 places and 23 transitions and it represents the interactions between T cells circulating in the peripheral blood and the blood–brain barrier (BBB), the barrier that divides blood from the central nervous system (CNS). Differently, with respect to [6], NK cells, effector memory T cells, and cytokines modules have been explicitly modeled and the CNS module has been replaced by just considering Regulatory Treg and Teff cells in CNS, and the ratio between them used as an indirect measure of CNS damage. Observe that Teff cells explicitly considered are Th1 and Th17 cells only, which share all the functions. As occurs in the pathological mechanism of MS, the cells in the blood vessels of the immune system can cross the blood–brain barrier (BBB) and reach the CNS and damage it. In this model Th17 cells have the function of increasing BBB permeability, reproducing this well-known biological effect of Th17 cells [30]. Also, the pathogen infection and then the reactivation in the system has been simulated differently from what was previously done. Recent data highlights Epstein-Barr virus (EBV) infection as a cause of MS, although is likely to be necessary, but not sufficient, to trigger its development [31]. Diverse possible mechanisms of EBV-mediated MS development have been proposed, including molecular mimicry, and B cell transformation [32]. EBV could contribute to MS through the reprogramming of latently infected B cells and drive a chronic presentation of viral antigens as a potential source of autoreactivity through molecular mimicry [33]. This chronic presentation of the antigen has been reproduced in the model using a sinusoidal function of antigen presentation through time. Finally, we reproduced alemtuzumab administration and all its connected functions: injection, degradation, consumption, and killing to reproduce the effect of the therapy on depleting all the mature circulating T cells.

3.4. From meta-patients to the generation of personalized therapy

To study model evolution, we first need to determine the value of the model parameters. Since the mechanism(s) of EBV-mediated MS development is difficult to determine [32], the antigen is modeled as an oscillatory function. For what concerns the therapy, it is simulated as two injections at baseline and at M12 following the clinical practice guidelines. All the other parameters are calibrated starting from the median dynamics of the immunological cell populations grouped in the $MSMP\_1$. In this regard, the $MSMP\_1$ dynamics are to be interpreted as the baseline.

The three columns of Figure C.13 report the dynamics obtained simulating the model exclusively changing the initial concentration of $CD4^+$, Th17, Th1, and Treg populations using the median values of $MSMP\_1$, $MSMP\_2$, and $MSMP\_3$, respectively.

In $MSMP\_2$ an increase in Th17 concentration is observed around 36 months. Fig. 3 (bottom row) reports the *whatif* analysis of the change in Th17 dynamics was obtained by varying the concentration of the antigen, modifying its oscillatory function in order to obtain a higher concentration peak around month 30. We simulated an additional infusion of alemtuzumab at month 36 and observed a change in T cell repopulation dynamics. It is interesting to note that after the additional alemtuzumab administration Th17, Th1, and Treg counts remain at lower levels compared to the values measured in $MSMP\_1$ and $MSMP\_3$, Fig. 3.

4. Discussion

Data from 6-year follow-up clinical trials show that alemtuzumab has durable efficacy in around one-half or less of the patients [21,22]. The remaining half of patients discontinued alemtuzumab treatment due to the inefficacy of therapy or to the manifestation of adverse

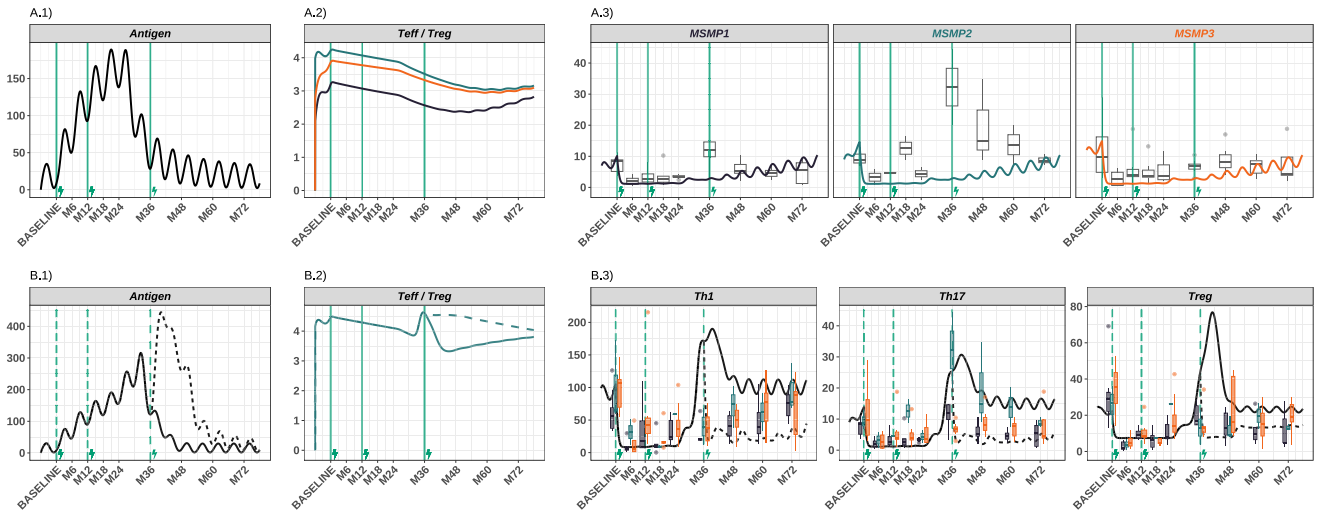

Fig. 3. In $A.1$ is reported the plot of the function used to model the antigen over time. In $A.2$ the Teff/Treg ratio of the three MSMP is shown revealing the clear difference in the first time points of the experiments between MSMP1 and the other MSMP groups. In $A.3$, are reported the predicted Th17 dynamics of all the MSMP groups. In $B.1$ the dynamic of the antigen is modified accordingly to reach the maximum value before M30 (continuous line). Moreover, the dynamic after an infusion of alemtuzumab at M36 is reported as a dashed line. In $B.2$ the Teff/Treg ratio plot is reported, the continuous line is associated with the change of antigen concentration at M30 while the dashed line describes the infusion of alemtuzumab. Finally, $B.3$ presents three plots of the simulation of Th1, Th17, and Treg cell populations, respectively. The $MSMP\_1$ is associated with dark blue color, $MSMP\_2$ is associated with teal color, and $MSMP\_3$ is associated with orange color. In all the plots, the continuous green vertical lines correspond to the infusion of alemtuzumab. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

events caused by deep and long-lasting immune suppression, including the development of secondary autoimmunity or opportunistic infections. A portion of patients with early relapse following initiation of alemtuzumab exists. Early relapsers could have an initial suboptimal response and the full immune modulatory effect may not be realized until the full course regimen. Results of a post hoc analysis showed that the outcome of the early relapsers improved after completing the second alemtuzumab course, supporting the fact that the administration of the approved two-course regimen should maximize the clinical benefit [34]. Our strategy of analysis using CONNECTOR and the unsupervised clustering algorithm, led to the stratification of alemtuzumab-treated patients into three MSMP groups, based on T-cell repopulation dynamics. MSMP_1 and MSMP_3 clusters contain MS patients with sparse relapse events, i.e. events that were non-repeated and occurred early in therapy or some years apart from alemtuzumab administration, and MSMP_2 cluster contains only patients with at least two relapse events that occurred early in the therapy and later after treatment. According to what was previously discussed, our analytic strategy efficiently clustered patients with similar T-cell dynamics that in the end have also similar clinical outcomes. A clear separation between "non-responders patients" in MSMP_2 cluster versus "responders" and "suboptimal responders" (i.e., patients with no relapses and patients that had a relapse but that later have clinical benefits from the second administration of the drug respectively) in clusters MSMP_3 and MSMP_1 is reported.

Modeling analysis of T-cell dynamics of the MSMP_2 cluster resulted in an increased Th17/Treg cells ratio in the CNS, with respect to MSMP_3 and MSMP_1. In our model, this result is an indirect measure of inflammation that drives neuronal damage. In the past years, immunological analyses on T-cell subsets did not succeed in capturing differences between MS patients with or without disease activity after alemtuzumab treatment [35,36].

More recently Rolla et al. [20] reported a difference in a higher Th17/Treg ratio at baseline leads to a higher relapse rate after treatment, this is also highlighted by our model analysis in which a change in the Th17/Treg ratio at baseline is reflected in a different T-cell repopulation trajectory.

The T-cell repopulation following alemtuzumab treatment may account for the need for an additional infusion in individual patients, but currently, available data have not identified a corresponding cellular biomarker. Real clinical data demonstrated that 1 or 2 additional injections after the first two doses effectively reduced the amount of post-treatment relapses, also reducing the worsening of disability; however, this study displayed some limitations due to the lack of a comparative group of patients who were eligible for additional alemtuzumab courses but did not receive them [37]. Therefore, it is still difficult to predict a priori whether a patient really needs further doses or not. In this study we simulated an additional treatment in patients of the MSMP_2 cluster, obtaining a dynamic of T-cell repopulation more similar to the other two clusters, suggesting that MSMP_2 could benefit from additional treatments. In this work, we used GreatNector on a limited number of subjects, and for a single type of treatment, of which we collected the necessary longitudinal data. To achieve an ideal patient stratification, the population of MS patients should be expanded through a longitudinal multicenter study involving several hospitals and institutions, in order to generalize the information found in this article. Ideal patient stratification would be further improved by considering different types of therapies, and even more so by considering the escalation of different treatments that often happens in clinical practice. Many of the currently approved therapies for MS consist of immunosuppressant drugs with diverse mechanisms of action (e.g., cladribine, ocrelizumab, ofatumumab) and therefore could be investigated with GreatNector. Specifically, gathering data from MS patients for the treatments of interest, we will utilize GreatNector to forecast the disease's progression and determine the optimal timing for administering the drug treatment.

The possibility of exploiting MS meta-patients to characterize the outcome of a patient from his/her baseline suggests that this approach could efficacy support physicians during the identification of patients that will need additional treatment before starting therapy by simulating patient-tailored interventions, based on meta-patient clustering.

Moreover, researchers and physicians will be able to exploit such MS meta-patients for investigating different intervention hypotheses aimed at contrasting the progression of the disease.

GreatNector and the analytic strategy proposed in this paper could be generalizable to any type of biological phenomena under study. The collection of longitudinal data is an experimental design increasingly common, while the challenging part is the definition of the mechanistic model to describe the systems under study in terms of interaction and

parameters. The model should be simple but not simplistic in order to capture the interactions among cells or molecules to explore the hypothesis that conducted the research.

5. Conclusion

GreatNector is a generalizable workflow designed by the combination of longitudinal data analysis and a quantitative modeling framework. In this paper, GreatNector is used to exploit the longitudinal data of immunological cell populations collected from 21 MS patients. The first part of GreatNector leads to define in an unsupervised manner three clusters able to distinguish groups characterized by patients with more than one relapse and patients without relapses. In the second part of GreatNector, the cluster information was integrated and exploited to calibrate the unknown parameters of a modified version of the MS model proposed in [6]. In detail, for each cluster of patients, we identified a specific parameter configuration, which was used (1) to simulate the model, capturing the variability characterizing the clusters, and (2) to study the individual IS response to alemtuzumab treatment. Therefore, in an era where therapies are focused on personalization, GreatNector aims to predict the evolution of MS and the response to treatments in specific patients, based on similar immunological structure at the time of the onset of the disease or therapy. Indeed, this workflow can lead to significant benefits for patients and improvements in their life conditions, that can be applied to other therapies different from the alemtuzumab treatment. Moreover, optimized therapies also mean more cost-effective therapies, since unnecessary and ineffective treatments will be reduced, with positive economic consequences.

CRediT authorship contribution statement

Simone Pernice: Conceptualization, Methodology, Software, Formal analysis, Writing — original draft, Visualization. **Alessandro Maglione:** Conceptualization, Data curation, Writing — original draft, Visualization. **Dora Tortarolo:** Data curation, Writing – original draft, Visualization. **Roberta Sirovich:** Methodology, Formal analysis, Writing – review & editing. **Marinella Clerico:** Resources, Funding acquisition, Writing – review & editing. **Simona Rolla:** Resources, Funding acquisition, Writing – original draft. **Marco Beccuti:** Conceptualization, Methodology, Formal analysis, Funding acquisition, Writing – original draft, Supervision. **Francesca Cordero:** Conceptualization, Methodology, Formal analysis, Funding acquisition, Writing – original draft, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Francesca Cordero reports financial support was provided by Horizon Europe. Marco Beccuti reports financial support was provided by CRT Foundation.

Availability

The two tools, i.e., CONNECTOR and GreatMod, which are exploited in this work are freely available at https://qbioturin.github.io/connector and https://qbioturin.github.io/epimod, respectively. The model, the parameters configuration and the file necessary to reproduce the results are available at https://github.com/qBioTurin/GreatNector-MS.

Acknowledgment

This research was supported by the scientific activities of the CINI InfoLife Laboratory.

Funding

CRT foundation *Creation of a computational framework to model and study West Nile Disease* [No 67410]. PRIN 2022 project *MEDICA: Modelling and vErification of alkaptonuria and multiple sclerosis Driven by biomedICAl data* [No 2022RNTYWZ].

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbi.2023.104546.

References

[1] C.B. Collin, T. Gebhardt, M. Golebiewski, T. Karaderi, M. Hillemanns, F.M. Khan, A. Salehzadeh-Yazdi, M. Kirschner, S. Krobitsch, Consortium, L. Kuepfer, Computational models for clinical applications in personalized medicine—guidelines and recommendations for data integration and model validation, J. Pers. Med. 12 (2) (2022) http://dx.doi.org/10.3390/jpm12020166.

[2] C.A. Petri, Kommunikation Mit Automaten., Insti- tut fur lnstrumentelle Mathematik, Schriften des IIM Nr. 3, 1962, Bonn, 1966.

[3] F. Mansoori, M. Rahgozar, K. Kavousi, A pathway analysis approach using Petri net, IEEE J. Biomed. Health Inf. 25 (3) (2021) 874–880, http://dx.doi.org/10.1109/JBHI.2020.3003996.

[4] J. Ashraf, J. Ahmad, A. Ali, Z. Ul-Haq, Analyzing the behavior of neuronal pathways in Alzheimer's disease using Petri net modeling approach, Front. Neuroinform. 12 (2018) http://dx.doi.org/10.3389/fninf.2018.00026.

[5] A. Kozak, D. Formanowicz, P. Formanowicz, Structural analysis of a Petri net model of oxidative stress in atherosclerosis, IET Syst. Biol. 12 (3) (2018) 108–117, http://dx.doi.org/10.1049/iet-syb.2017.0015.

[6] S. Pernice, L. Follia, A. Maglione, M. Pennisi, F. Pappalardo, F. Novelli, M. Clerico, M. Beccuti, F. Cordero, S. Rolla, Computational modeling of the immune response in Multiple Sclerosis using Epimod framework, BMC Bioinf. 21 (17) (2020) 1–20.

[7] L. Peng, P. Xie, Z. Tang, F. Liu, Modeling and analyzing transmission of infectious diseases using generalized stochastic Petri nets, Appl. Sci. 11 (18) (2021) http://dx.doi.org/10.3390/app11188400, URL https://www.mdpi.com/2076-3417/11/18/8400.

[8] P. Castagno, S. Pernice, G. Ghetti, M. Povero, L. Pradelli, D. Paolotti, G. Balbo, M. Sereno, M. Beccuti, A computational framework for modeling and studying pertussis epidemiology and vaccination, BMC Bioinf. 21 (2020) 344, http://dx.doi.org/10.1186/s12859-020-03648-6.

[9] A.R. Udyavar, D.J. Wooten, M. Hoeksema, M. Bansal, A. Califano, L. Estrada, S. Schnell, J.M. Irish, P.P. Massion, V. Quaranta, Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity, Cancer Res. 5 (77) (2017) 1063–1074.

[10] F. Eduati, P. Jaaks, J. Wappler, T. Cramer, C.A. Merten, M.J. Garnett, J. Saez-Rodriguez, Patient-specific logic models of signaling pathways from screenings on cancer biopsies to prioritize personalized combination therapies, Mol. Syst. Biol. 16 (2) (2020) e8664.

[11] F.M. Khan, S. Marquardt, S.K. Gupta, S. Knoll, U. Schmitz, A. Spitschak, D. Engelmann, J. Vera, O. Wolkenhauer, B.M. Pützer, Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures, Nature Commun. 8 (1) (2017) 198, http://dx.doi.org/10.1038/s41467-017-00268-2.

[12] G. Weatherley, R.P. Araujo, S.J. Dando, A.L. Jenner, Could mathematics be the key to unlocking the mysteries of multiple sclerosis? Bull. Math. Biol. 85 (2023) 75–80, http://dx.doi.org/10.1007/s11538-023-01181-0.

[13] A. Maleki, E. Crispino, S.A. Italia, V. Di Salvatore, M.A. Chiacchio, F. Sips, R. Bursi, G. Russo, D. Maimone, F. Pappalardo, Moving forward through the in silico modeling of multiple sclerosis: Treatment layer implementation and validation, Comput. Struct. Biotechnol. J. 21 (2023) 3081–3090, http://dx.doi.org/10.1016/j.csbj.2023.05.020.

[14] S. Pernice, R. Sirovich, E. Grassi, M. Viviani, M. Ferri, F. Sassi, L. Alessandrí, D. Tortarolo, R.A. Calogero, L. Trusolino, A. Bertotti, M. Beccuti, M. Olivero, F. Cordero, CONNECTOR, fitting and clustering of longitudinal data to reveal a new risk stratification system, Bioinformatics 39 (5) (2023) btad201, http://dx.doi.org/10.1093/bioinformatics/btad201.

[15] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer Science & Business Media, 2006.

[16] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, 2005.

[17] S. Pernice, M. Pennisi, G. Romano, A. Maglione, S. Cutrupi, F. Pappalardo, G. Balbo, M. Beccuti, F. Cordero, R.A. Calogero, A computational approach based on the Colored Petri Net formalism for studying Multiple Sclerosis, BMC Bioinf. (2019).

[18] S. Pernice, M. Beccuti, G. Romano, M. Pennisi, A. Maglione, S. Cutrupi, F. Pappalardo, L. Capra, G. Franceschinis, M. De Pierro, G. Balbo, F. Cordero, R. Calogero, Multiple sclerosis disease: A computational approach for investigating its drug interactions, Lecture Notes in Comput. Sci. 12313 LNBI (2020) 299–308.
[19] G. Giovannoni, Personalized medicine in multiple sclerosis, Neurodegener. Dis. Manag. 7 (6s) (2017) 13–17.
[20] S. Rolla, S.F. De Mercanti, V. Bardina, A. Maglione, D. Taverna, F. Novelli, E. Cocco, A. Vladic, M. Habek, I. Adamec, et al., Long-term effects of alemtuzumab on CD4+ lymphocytes in multiple sclerosis patients: A 72-month follow-up, Front. Immunol. 13 (2022).
[21] A.J. Coles, J.A. Cohen, E.J. Fox, G. Giovannoni, H.-P. Hartung, E. Havrdova, S. Schippling, K.W. Selmaj, A. Traboulsee, D.A.S. Compston, et al., Alemtuzumab CARE-MS ii 5-year follow-up: efficacy and safety findings, Neurology 89 (11) (2017) 1117–1126.
[22] B. Steingo, Y. Al Malik, A.D. Bass, R. Berkovich, M. Carraro, Ó. Fernández, C. Ionete, L. Massacesi, S.G. Meuth, D.D. Mitsikostas, et al., Long-term efficacy and safety of alemtuzumab in patients with RRMS: 12-year follow-up of CAMMS223, J. Neurol. 267 (2020) 3343–3353.
[23] S. Wray, E. Havrdova, D.R. Snydman, D.L. Arnold, J.A. Cohen, A.J. Coles, H.-P. Hartung, K.W. Selmaj, H.L. Weiner, N. Daizadeh, et al., Infection risk with alemtuzumab decreases over time: pooled analysis of 6-year data from the CAMMS223, CARE-MS I, and CARE-MS II studies and the CAMMS03409 extension study, Multiple Scler. J. 25 (12) (2019) 1605–1617.
[24] S. Meyer-Moock, Y.-S. Feng, M. Maeurer, F.-W. Dippel, T. Kohlmann, Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis, BMC Neurol. 14 (1) (2014) 1–10.
[25] S. Rolla, A. Maglione, S.F. De Mercanti, M. Clerico, The meaning of immune reconstitution after alemtuzumab therapy in multiple sclerosis, Cells 9 (6) (2020) 1396.
[26] G.M. James, C.A. Sugar, Clustering for sparsely sampled functional data, J. Amer. Statist. Assoc. 98 (462) (2003) 397–408.
[27] G.M. James, T.J. Hastie, C.A. Sugar, Principal component models for sparse functional data, Biometrika 87 (3) (2000) 587–602.
[28] E.G. Amparore, G. Balbo, M. Beccuti, S. Donatelli, G. Franceschinis, 30 Years of GreatSPN, in: Principles of Performance and Reliability Modeling and Evaluation, Springer, 2016, pp. 227–254.
[29] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2023, URL https://www.R-project.org/.
[30] H. Kebir, K. Kreymborg, I. Ifergan, A. Dodelet-Devillers, R. Cayrol, M. Bernard, F. Giuliani, N. Arbour, B. Becher, A. Prat, Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation, Nat. Med. 13 (10) (2007) 1173–1175.
[31] K. Bjornevik, M. Cortese, B.C. Healy, J. Kuhle, M.J. Mina, Y. Leng, S.J. Elledge, D.W. Niebuhr, A.I. Scher, K.L. Munger, et al., Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis, Science 375 (6578) (2022) 296–301.
[32] W.H. Robinson, L. Steinman, Epstein-Barr virus and multiple sclerosis, Science 375 (6578) (2022) 264–265.
[33] S.S. Soldan, P.M. Lieberman, Epstein–Barr virus and multiple sclerosis, Nat. Rev. Microbiol. 21 (1) (2023) 51–64.
[34] B. Van Wijmeersch, B.A. Singer, A. Boster, S. Broadley, Ó. Fernández, M.S. Freedman, G. Izquierdo, J. Lycke, C. Pozzilli, B. Sharrack, et al., Efficacy of alemtuzumab over 6 years in relapsing–remitting multiple sclerosis patients who relapsed between courses 1 and 2: Post hoc analysis of the CARE-MS studies, Multiple Scler. J. 26 (13) (2020) 1719–1728.
[35] O. Kousin-Ezewu, L. Azzopardi, R.A. Parker, O. Tuohy, A. Compston, A. Coles, J. Jones, Accelerated lymphocyte recovery after alemtuzumab does not predict multiple sclerosis activity, Neurology 82 (24) (2014) 2158–2164.
[36] H. Wiendl, M. Carraro, G. Comi, G. Izquierdo, H.J. Kim, B. Sharrack, C. Tornatore, N. Daizadeh, L. Chung, A.K. Jacobs, et al., Lymphocyte pharmacodynamics are not associated with autoimmunity or efficacy after alemtuzumab, Neurol. Neuroimmunol. Neuroinflammation 7 (1) (2020).
[37] G. Comi, R. Alroughani, A.L. Boster, A.D. Bass, R. Berkovich, Ó. Fernández, H.J. Kim, V. Limmroth, J. Lycke, R.A. Macdonell, et al., Efficacy of alemtuzumab in relapsing-remitting MS patients who received additional courses after the initial two courses: pooled analysis of the CARE-MS, extension, and TOPAZ studies, Multiple Scler. J. 26 (14) (2020) 1866–1876.