Computers in Biology and Medicine 50 (2014) 19–31

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Time-series analysis in the medical domain: A study of Tacrolimus administration and influence on kidney graft function

Vladimir Kurbalija [a,*], Miloš Radovanović [a], Mirjana Ivanović [a], Danilo Schmidt [b], Gabriela Lindemann von Trzebiatowski [c], Hans-Dieter Burkhard [c], Carl Hinrichs [d]

[a] University of Novi Sad, Faculty of Science, Department of Mathematics and Informatics, Trg D. Obradovića 4, 21000 Novi Sad, Serbia
[b] Charité – Universitätsmedizin Berlin, Department of Nephrology, Campus Charité Mitte, Germany
[c] Humboldt University of Berlin, Department of Computer Science, Germany
[d] Charité – Universitätsmedizin Berlin, Department of Nephrology, Campus Virchow-Hospital, Germany

ARTICLE INFO

Article history:
Received 3 December 2013
Accepted 11 April 2014

Keywords:
Kidney transplantation
Tacrolimus
Creatinine
Glomerular filtration rate
Time series
Distance measures

ABSTRACT

There exists a major concern regarding toxic effects of immunosuppressive medication on the kidney graft during post-transplant care, with observed variation in individual susceptibility to adverse drug effects amongst patients. To date, there has been no possibility to identify susceptible patients prospectively. This study analyzes medical data which includes time series of measures of renal function and trough levels of immunosuppressive drug Tacrolimus, with the main aim of identifying patients susceptible to drug toxicity. We evaluate a plethora of time-series distance measures, determining their appropriateness to the domain based on two criteria: (1) preserving the expected correlations between distances, and (2) ability to detect the expected patterns of interaction between immunosuppressive drug levels and renal function. Besides identifying the most suitable time-series distance measures, we observed that the majority of patients do not exhibit an association between impaired graft function and higher Tacrolimus dosing. On the other hand, the minority of patients determined most sensitive to varying Tacrolimus levels showed a strong tendency to prefer low Tacrolimus dosing.

1. Introduction

End stage renal disease (ESRD) is one of the major health problems in western industrialized countries and is one of the main drivers of health-care related costs. Due to demographic changes in industrialized and transforming countries, its incidence is growing annually by around 5–8%, causing more than 2.5 billion Euros of health-care related costs in Germany. While dialysis is the standard treatment for end-stage renal disease, renal transplantation is an alternative which is superior regarding survival, life quality and health-care related costs. However, its availability is reduced due to legal issues that cause a scarcity of suitable donor organs. Thus, post-transplant care aims at stabilizing kidney function as long as possible to reduce the need for re-transplantation. While in the early phase of transplantation, graft survival was mainly determined by the occurrence of immunological adverse events (such as “acute rejection”), to date, a major concern are toxic effects of immunosuppressive medication on the kidney graft.

A wide variety of complex data is produced during the post-transplant care. Besides demographic data which do not change over time, most of the data consists of time-series, as different examinations are undertaken on a patient repetitively. Our medical data come from the Department of Nephrology of the University Hospital Charité at Berlin and include time-series of measures of renal function and trough levels of immunosuppressive medication [1].

Clinical experience shows that individual susceptibility to adverse drug effects varies amongst patients. While some react to an increased exposure to immunosuppressive medication, with an imm ediate drop in renal function, others have stable graft function over an ample range of drug intake. To date, there is no possibility to identify susceptible patients prospectively.

In the study presented here, we evaluate different similarity measures to compare time-series of drug trough levels and renal function. We aim at identifying patients who are susceptible to drug toxicity. Thereby, we would like to identify candidate patients that might benefit from a switch to less nephrotoxic drug regimens. To achieve this goal we evaluate a plethora of time-series distance measures and determine their appropriateness to the domain based on two criteria: preserving the expected correlations between

* Corresponding author. Tel.: +381 21 4852877.
E-mail addresses: kurba@dmi.uns.ac.rs (V. Kurbalija), radacha@dmi.uns.ac.rs (M. Radovanović), mira@dmi.uns.ac.rs (M. Ivanović), danilo.schmidt@charite.de (D. Schmidt), gabriela.lindemann@uv.hu-berlin.de (G.L. vonTrzebiatowski), hdb@informatik.hu-berlin.de (H.-D. Burkhard), carl.hinrichs@charite.de (C. Hinrichs).

http://dx.doi.org/10.1016/j.compbiomed.2014.04.007
0010-4825/

distances in the data (Section 3.1), and ability to detect the expected patterns of interaction between immunosuppressive drug levels and renal function (Section 3.2).

1.1. Medical background

End stage renal disease is the notion of irreversible loss of renal function and denotes the common final state of most severe kidney diseases. It is fatal if not treated by renal replacement therapy. Today, two basic options exist for renal replacement, namely dialysis and renal transplantation. While renal transplantation is far superior to dialysis regarding quality of life and patient's survival (LIT), its main drawback is the need for constant intake of medication that suppress immunological reaction against the donor kidney, the so called graft rejection. This medication has a narrow therapeutic index and some serious adverse effects, of which the most relevant are the increased risk of infection and cancer, metabolic disorders and chronic kidney damage. Therefore, therapy is guided by determination of trough levels, which means measurement of the plasma concentrations of the respective substance in the morning before intake of a new dose.

For the therapy control, monitoring of renal function is necessary. The gold standard for the measurement of renal function is the determination of glomerular filtration rate (GFR), a measure reflecting the volume of plasma which is cleared from a substance per minute. In clinical routine the plasma concentration of creatinine is used as a surrogate for GFR measurement. Creatinine is an end product of muscle metabolism and is filtrated by the kidney at a high ratio. To avoid some drawbacks associated with the sole determination of creatinine, GFR is estimated based on plasma creatinine concentration, gender, age and ethnicity of patients using the Modification of Diet in Renal Disease formula (named estimated GFR, eGFR).

Clinical experience suggests that susceptibility to kidney damage by immunosuppressive drugs varies from patient to patient. While some patients react to a slight overdose of immunosuppression with an immediate drop in eGFR, the tolerable range in other patients is larger. To avoid chronic (and irreversible) graft damage, the identification of high risk patients would allow taking countermeasures such as adaption of individual target trough levels or change of medication. Moreover, the systematic identification of patients with a high similarity of eGFR and trough level time courses would allow identifying demographic differences between more and less susceptible patients, thereby facilitating a suitable choice of immunosuppression at an early time after transplantation.

1.2. Time series

Time series represent a form of data where the value of some numeric parameter in an observed process is recorded periodically in time. The parameters of interest can be various: the recorded values of some scientific sensor, the daily number of hits of some website, the value of shares on a stock market, the number of rainy days per year, etc. The observation rate can vary from several milliseconds to several years, and in the general case the observation rate need not to even be uniform. However, the common characteristic of time series is that they usually consist of a large number of points (recorded values), which can make the handling of time series difficult.

Time-series analysis applies different methods, mainly from statistics, data mining and machine learning, in order to model and understand the process which generated time series or to make forecasts. Accordingly, several task types are important for time-series analysis, most notably indexing, classification, clustering, prediction, segmentation, and anomaly detection [2–5]. In addition, there are three important concepts which need to be considered when dealing with time-series data:

- Pre-processing transformation (the preparation of time series for further analysis and removing different kinds of distortion, noise and linear trend);
- Time-series representation (reducing the number of points in time series making them easier to handle, but keeping the overall shape or some other property of time series);
- Similarity/distance measure (similarity is crucial for all mentioned task types; the similarity/distance between time series should be carefully defined in order to reflect the underlying similarity or dissimilarity of these specific data which is usually based on shapes and patterns).

In this article we focus primarily on the choice of an appropriate similarity/distance measure for the task of comparing immunosuppressive drug trough levels and renal function indicators.

1.3. Time-series similarity measures

A similarity or distance measure is a crucial part of many time-series analysis applications. Such a measure is defined as a function which expresses to what extent the two time series are similar or different. However, this measure should be very carefully defined in order to reflect the semantically appropriate similarity or dissimilarity usually based on overall shapes of time series.

The most intuitive distance measure is the L_p norm defined for time series $Q=(q_1, q_2, \ldots, q_n)$ and $C=(c_1, c_2, \ldots, c_n)$, for an arbitrary positive rational number p:

$$L_p(Q,C)=\sqrt[p]{\sum_{i=1}^{n}|q_i-c_i|^p} \tag{1}$$

For different values of p, different distance measures can be obtained: $p=1$ – Manhattan (city-block) distance; $p=2$ – Euclidean distance; $p=\infty$ – Chebyshev (maximum, supremum, “sup”) distance, $p=1/2$ – fractional $L_{1/2}$ distance, etc. Here, the distance between two time series is calculated as the p-root of the sum of distances between corresponding points of the series taken to the pth power, with the distances between corresponding points illustrated in Fig. 1A. The major advantages of L_p distances are its computational efficiency and intuitive clarity which resulted in a large amount on applications [6–9]. However, some serious disadvantages of this measure are the sensitivity to distortions and shifting along the time axis, and the necessity of comparing time series of equal length.

Some of these disadvantages are corrected with the Dynamic Time Warping (DTW) measure. This measure allows non-linear aligning of the points of time series [10–13] as illustrated by the example in Fig. 1B.

Another important measure is the Edit distance with Real Penalty (ERP) based on the L_1 and DTW measures. It solves the problem of local time shifting and represents a metric [14]. ERP introduces a constant value g (with the default value $g=0$) as the gap of edit distance and uses L_1 distance between elements as the penalty to handle local time shifting.

$$D(i,j)=\begin{cases}0 & i=j=0\\ \infty & i=0, j>0 \vee i>0, j=0\\ d(q_i,c_j)+\min\begin{cases}D(i-1,j-1)\\ D(i-1,j)\\ D(i,j-1)\end{cases} & i,j\geq 1\end{cases} \tag{2}$$