Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–22

DOI: 10.1177/0962280215598805
smm.sagepub.com
SAGE

Applications of temporal kernel canonical correlation analysis in adherence studies

Majnu John,[1,2,3] Todd Lencz,[1,2,4] Janina Ferbinteanu,[5] Juan A Gallego[1,2,4] and Delbert G Robinson[1,2,4]

Abstract

Adherence to medication is often measured as a continuous outcome but analyzed as a dichotomous outcome due to lack of appropriate tools. In this paper, we illustrate the use of the temporal kernel canonical correlation analysis (tkCCA) as a method to analyze adherence measurements and symptom levels on a continuous scale. The tkCCA is a novel method developed for studying the relationship between neural signals and hemodynamic response detected by functional MRI during spontaneous activity. Although the tkCCA is a powerful tool, it has not been utilized outside the application that it was originally developed for. In this paper, we simulate time series of symptoms and adherence levels for patients with a hypothetical brain disorder and show how the tkCCA can be used to understand the relationship between them. We also examine, via simulations, the behavior of the tkCCA under various missing value mechanisms and imputation methods. Finally, we apply the tkCCA to a real data example of psychotic symptoms and adherence levels obtained from a study based on subjects with a first episode of schizophrenia, schizophreniform or schizoaffective disorder.

Keywords

Adherence, canonical correlation analysis, kernel methods, time series

1 Introduction

One of the most challenging aspects of health care is patient adherence (previously designated compliance) to medication, the extent to which patients take medications as prescribed by their health care providers.[1] Numerous studies have shown that in addition to the personal suffering caused by relapse of disease, premature disability, or death, poor adherence also leads to

[1]Center for Psychiatric Neuroscience, The Feinstein Institute of Medical Research, Manhasset, NY, USA
[2]Psychiatry Research, Zucker Hillside Hospital, North Shore LIJ Health System, Glen Oaks, NY, USA
[3]Department of Mathematics, Hofstra University, Hempstead, NY, USA
[4]Departments of Psychiatry and of Molecular Medicine, Hofstra North Shore LIJ School of Medicine, Hempstead, NY, USA
[5]Department of Physiology and Pharmacology, State University of New York, Health Science Center at Brooklyn, Brooklyn, NY, USA

Corresponding author:
Majnu John, Center for Psychiatric Neuroscience, The Feinstein Institute of Medical Research, Manhasset 11030, NY, USA.
Email: mjohn5@nshs.edu

increased health care costs.[2–8] As reported by Osterberg and Blaschke,[1] 33–69% of all medical related hospital admissions in the USA are due to poor adherence, with a resultant cost of approximately $100 billion a year.[2,4] These important clinical and societal costs have driven development of adherence research as a focus of investigation across multiple branches of medicine. Even for studies not directly focused upon adherence, understanding the effects of poor adherence to a study medication can be crucial for interpreting study outcomes.

Adherence in studies is often analyzed as a dichotomous variable (adherence vs. non-adherence) or as a percent of medication taken over defined periods. These methods can model well patient behavior for patients with sustained periods of similar levels of adherence (e.g. sustained periods of adherence or total refusal of medication). However, patient adherence behavior is often more complex. Patients may have intermittent non-adherence (e.g. taking medications one day and not the next) or partial adherence consisting of taking some but not all of their prescribed medication. Methods that quantify adherence as a dichotomous variable or percent medication taken do not model well complex non-adherence behaviors and may therefore fail to detect important outcome differences. For example, patients who take half their prescribed dose daily, patients who take their full dose every other day and none other days, and patients who take their prescribed dose daily for half a period but none afterward could be assigned the same percent adherence but have different outcomes. In the past, obtaining the data needed for modeling complex medication behaviors was challenging. Pill counts, for example, can only provide estimates of adherence averaged over the period between counts. However, the development of electronic medication monitors such as chips that record the data and the time of every pill ingestion provide methods to obtain data on adherence on a daily, or even more frequent, basis. The ability to obtain these richer data sets suggests the need to consider whether our analytic approaches may need to be expanded.

Traditional analytic strategies also often focus upon outcomes entered as dichotomous outcome, for example, relapse of a disease. Usually, these dichotomous outcomes are obtained by using a cut-off point on a continuous scale of symptoms.[7,8] Even if the cut-offs are based on clinically meaningful definitions, in general, it is more informative and efficient to analyze the continuous scale without dichotomization.[9]

This paper presents a method to analyze the effect of non-adherence on symptom levels using the whole continuum of both adherence and symptom measurements. The data are considered as two time series, one affecting the other and with a lag between them. It is often of interest to estimate the lag (on the average across all the patients) between the two series. For example, one may think of psychotic symptoms after first response of a patient with schizophrenia and adherence measured as dose levels of a treatment drug (e.g. risperidone) as two time series. As the adherence levels drop off occasionally, the psychotic symptoms return to pre-response levels. This in turn may bring the adherence levels back to the response dose level causing the psychotic symptoms to decrease or disappear again after a lag. This is an example of two time series mutually affecting or interacting with each other for one patient. Observed on multiple patients, we have multiple time series of psychotic symptoms and multiple time series of adherence levels. Researchers involved in adherence studies of mental health patients are often interested in knowing the average time lag. In the above example, the highest correlation between the two time series is obtained at some time shift of one of the series. The main question of interest is to determine the time shift at which the highest correlation occurs.

The two main issues to be dealt with when working with such questions are (*a*) the inherent multivariate nature of the data (each time series measured on multiple subjects) and (*b*) noninstantenous coupling between the two time series. If the two times series were coupled instantaneously (that is, no lag), the multivariate nature of the problem could have been handled

by canonical correlation analysis (CCA) which finds linear combinations of cross sections of the two series that maximizes the correlation between them. Specifically, CCA first determines the pair of linear combinations (canonical variables) having the largest correlation, and then recursively at the *k-th* step determines the pair of linear combinations which maximizes the correlation among all choices uncorrelated with the previous k-1 canonical variables.[10] If the number of dimensions is very large, it becomes more efficient and numerically more stable to optimize the linear combinations using kernel methods, i.e. kernel CCA.[11–13] In the presence of non-instantaneous coupling, the above methods are insufficient; however, in this case, a recently developed approach called the temporal kernel canonical correlation analysis (tkCCA) can be utilized.[14,15] The tkCCA takes into account the multivariateness as well as the temporal dynamics of the coupling by looking at canonical correlations at different time lags. Maximization is performed as a whole, distinguishing it from standard correlation analysis which simply computes the similarities between the two series individually.

The original application that led to the development of the tkCCA was to examine the relationship between neural signal fluctuations and blood oxygen level-dependent (BOLD) contrast obtained by functional magnetic resonance imaging (fMRI) during spontaneous activity.[15] But, the tkCCA is a technique that can have wide applications outside the area of its original development. In this paper, we illustrate how the tkCCA can be applied to adherence studies. First, we give a notational introduction to the tkCCA. Then we illustrate the use of the tkCCA with simulations and an application to a real-world dataset. We conclude the paper with a discussion on the possible use of the tkCCA in adherence research in the future.

2 Temporal kernel canonical correlation analysis

The theory presented in this section was developed by Bieβmann and colleagues[14]; we summarize it here for the sake of completeness. Let us assume that $X \in R^{N\times T}$ and $Y \in R^{N\times T}$ denote two time series with T time points measured on N subjects. That is, ith row of X and Y respectively correspond to the two time series measured on the ith patient; for example, the series of psychotic symptoms and the series of adherence levels of a patient with schizophrenia. In the presence of instantaneous coupling, we could have used CCA on any cross section of the two time series, x_j and y_j, where x_j and y_j denote the jth columns of X and Y, respectively. We may consider x_j and y_j, $j = 1, \ldots, T$, as T realizations (samples) of two random variables x and y, respectively. For the sake of simplicity, we assume that x and y are centered; that is, the underlying means are zero. In the first step, CCA estimates two normalized vectors a_x and b_y in R^N (first pair of canonical coefficients) such that the correlation between the projections $\langle a_x, x\rangle$ and $\langle b_y, y\rangle$ is maximized[10]

$$\begin{aligned} &\underset{a_x,b_y}{\arg\max}\ \mathrm{Corr}(\langle a_x, x\rangle), (\langle b_y, y\rangle) \\ &= \underset{a_x,b_y}{\arg\max}\left(\frac{\mathrm{E}\big(\langle a_x, x\rangle\langle b_y, y\rangle\big)}{\sqrt{\mathrm{Var}(\langle a_x, x\rangle)\mathrm{Var}\big(\langle b_y, y\rangle\big)}}\right) \\ &\text{such that } \mathrm{Var}(\langle a_x, x\rangle) = 1 = \mathrm{Var}\big(\langle b_y, y\rangle\big) \end{aligned} \tag{1}$$

where $\langle\cdot,\cdot\rangle$, E and Var respectively denote the inner product, expectation and variance; we require unit variances to reduce the freedom of scaling of the projections. The optimal correlation in (1) is

called the first canonical correlation and corresponding projections that maximize the first canonical correlation are called the first canonical variables.[10] Note that a_x and b_y can also be obtained as an eigenvector corresponding to the maximal eigenvalues of a generalized eigenvalue problem. If x and y are variables of dimension 1, we stop at step 1. If the dimension (that is, N) is greater than 1, we would proceed recursively to identify further canonical variables. For example, the second pair of canonical variables is the pair of projections having unit variances, which maximize the correlation in (1) among all choices that are uncorrelated with the first pair of canonical variables, and so on. The kth canonical correlation is always greater than the $(k+1)^{st}$ canonical correlation, and the first canonical correlation is interpreted as a measure of overlap between the original set of variables.

For high dimensional multivariate data, the above approach may not be computationally optimal. In such cases, kernel-based methods have been suggested to obtain the canonical coefficients in (1). In addition to computational efficiency, kernel-based methods are also useful if the relationship between the two original set of variables is nonlinear. When the columns of X and Y are centered, the linear kernel matrices K_X and K_Y are given as the inner product of the data matrices

$$\begin{aligned} K_X &= X^T X, \\ K_Y &= Y^T Y. \end{aligned} \tag{2}$$

The canonical coefficients of each variable are then given as a linear expansion of the data points

$$\begin{aligned} a_x &= Xu, \\ b_y &= Yv \end{aligned} \tag{3}$$

where the $u \in R^T$ and $v \in R^T$ are the solutions of the generalized eigenvalue problem in kernel space

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \rho \begin{bmatrix} K_X^2 & 0 \\ 0 & K_Y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{4}$$

The above equation can be obtained by solving the Lagrangian corresponding to the optimization problem given in (1). For the data analysis in sections 3 and 4, we used the *geigen* package[16] in R to solve equation (4). For the sake of simplicity, we consider only the largest eigenvalue ρ (*i.e.* the first canonical correlation coefficient). In practice, we have to introduce regularization parameters β_X and β_Y in equation (4) to prevent overfitting

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \rho \begin{bmatrix} K_X^2 + \beta_X K_X & 0 \\ 0 & K_Y^2 + \beta_Y K_Y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{5}$$

In the discussion of behavioral health research presented in the introduction, we saw that the highest correlation is obtained at some time shift between the two time series. In general, we have to assume that there is not just a fixed delay, but that the canonical coefficients have a temporal dimension. We will therefore generalize the CCA optimization problem by allowing a_x to be time-dependent[14,15]

$$\underset{a_x(\tau), b_y}{\arg\max}\, \mathrm{Corr}\left(\sum_{\tau} \langle a_x(\tau), X_\tau \rangle, \langle b_y, Y \rangle \right). \tag{6}$$

We may rewrite the above optimization problem by letting

$$\tilde{X} = \begin{bmatrix} X_{\tau_1} \\ X_{\tau_2} \\ \vdots \\ X_{\tau_L} \end{bmatrix} \in R^{NL \times T} \text{ and } \tilde{a}_x = \begin{bmatrix} a_x(\tau_1) \\ a_x(\tau_2) \\ \vdots \\ a_x(\tau_L) \end{bmatrix} \in R^{NL \times T}, \tag{7}$$

so that it reduces to an ordinary CCA problem

$$\underset{\tilde{a}_x, b_y}{\arg\max}\ \mathrm{Corr}\big(\langle \tilde{a}_x, \tilde{X} \rangle, \langle b_y, Y \rangle\big). \tag{8}$$

Here, X_{τ_j}, $j = 1, \ldots, L$ are time-shifted copies of the data matrix X, each shifted in time by a lag from the set $\{\tau_1, \ldots, \tau_L\}$. It is by looking at canonical correlations at different time lags that the tkCCA accounts for the temporal dynamics of the coupling. With $\tilde{K}_X = \langle \tilde{X}, \tilde{X} \rangle$ instead of K_X in the generalized eigenvalue problem (5), we may obtain the canonical coefficients for the optimization problem in (8) as

$$\begin{aligned} \tilde{a}_x &= \tilde{X} u, \\ \tilde{b}_y &= \tilde{Y} v. \end{aligned} \tag{9}$$

In particular, we may recover the canonical coefficient for each time lag by

$$a_x(\tau) = X_\tau u. \tag{10}$$

The results of the tkCCA may be used to compute a canonical cross-correlogram similar to a standard univariate cross-correlogram

$$\begin{aligned} \rho(\tau) &= \mathrm{Corr}\big(\langle a_x(\tau), X_\tau \rangle, \langle b_y, Y \rangle\big) \\ &= \frac{\langle a_x(\tau), X_\tau \rangle \langle b_y, Y \rangle}{(\langle a_x(\tau), X_\tau \rangle)^2 (\langle b_y, Y \rangle)^2}. \end{aligned} \tag{11}$$

3 Simulated illustrative examples

In this section, we illustrate the above concepts using simulated data. In section 3.1, we generate univariate time series via simulation, and in section 3.2 we generate sets of multivariate times series to apply and illustrate the tkCCA method.

3.1 Univariate time series example

The effect of non-adherence may vary from patient to patient. Determining the time lag between non-adherence and return of symptoms for each individual patient can have important utility in a clinic. For example, a clinician may schedule more hospital visits for a patient who has a lag smaller than the average patient. A recently developed drug-delivery strategy for psychiatric patients is the extended or sustained release drug delivery, specifically based on long acting drug formulations.[17] In this strategy, instead of daily doses, the drug is administered alternate days, or weekly, for example.

The time interval is decided on a patient to patient basis. A patient with quicker relapse would be given more frequent doses compared to a patient with slower relapse so it is useful to know the time lag between adherence and symptom levels.

We consider a hypothetical symptom scale for a hypothetical brain disorder. We assumed that the symptom scale is based on four items, each ranging from 1 to 7, so that the total symptom score obtained by taking the sum of these four items ranges from 4 to 28. We generated symptoms score for a hypothetical patient for 100 weeks as an autoregressive integrated moving average (ARIMA) model, $p(t)$, with autoregressive (AR) order set to 1, and degree of differencing and moving average (MA) order set to 0, and AR coefficients set to 0.9 (Figure 1, top panel). The following R codes were used for this step

```
set.seed(54)
symps ← arima.sim(list(order = c(1, 0, 0), ar = .9), n = 100) + 10
```

The dose level for each patient can vary, but we assumed that the dose level at which the hypothetical patient in our illustration responded initially was around 3 mg. Note that typically the prescribed dose level may increase from the initial-response–dose level if the symptoms progressively worsen due to non-adherence. We generated the dose level time series, $d(t)$, as a scaled (by a factor of 0.33) version of the symptoms time series, $p(t)$, shifted to the right by two units (weeks) and with Gaussian white noise added to it:

$$d(t) = \frac{1}{3}p(t+2) + \frac{1}{2}N(t), \tag{12}$$

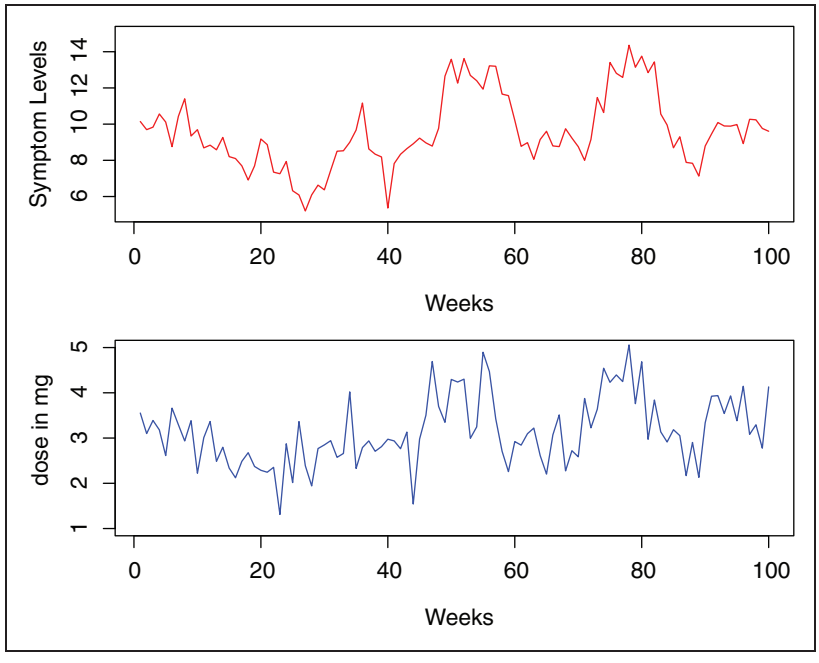

Figure 1. Simulated plot of symptoms and adherence levels for a patient (with a hypothetical brain disorder) with high fluctuations of adherence: Univariate time series example.

where $N(t) = N(0, 1)$ is Gaussian i.i.d. The codes used for this step are shown as follow:

```
d ← symps/3
set.seed(50)
dose ← c(d[3 : 100], d[1 : 2]) + 0.5 ∗ rnorm(100)
```

The univariate time series generated via these codes are shown in Figure 1, bottom panel, and reflects the fluctuations in adherence from the prescribed response-dose level. We, a priori, set the lag between the two simulated time series to be 2 units (weeks). That is, we assumed that it takes approximately two weeks for symptoms to worsen, when the patient does not adhere to the prescribed dose level. The goal of the analysis was to check whether the lag found by the tkCCA method would match the lag that was set a priori.

The lags that we considered for the tkCCA method, equations (7) to (10) are

$$-25, -24, -23, \ldots, 0, \ldots, 23, 24 \text{ and } 25 \text{ units.}$$

The lag set to be considered will typically depend on the clinical question. In general, we recommend to set the lag set as large as possible. The correlogram based on the absolute values of $\rho(\tau)$ in equation (11) is shown in Figure 2. The peak of the correlogram is at a lag of 2 weeks, and so, indeed the tkCCA method recovers the lag that was set a priori.

We conducted Monte Carlo simulations to see whether the performance of the tkCCA method depended on the lag or the variance of the noise. For each Monte Carlo iteration, a symptom series similar to the series in Figure 1 was generated by the following code:

```
set.seed(54)
symps1 ← arima.sim(list(order = c(1, 0, 0), ar = .9), n = 100) + 10
#In the Monte Carlo loop...
symps ← symps1 + 1.5 ∗ N(0, 1)
```

Note that the symptoms series varies at each iteration because of the $1.5 * N(0, 1)$ noise term added to it. As in the aforesaid example, at each iteration, we obtained the dose series, first by scaling the symptoms series by a factor of 1/3 and then adding a lag and noise term to it. We conducted 10,000 Monte Carlo iterations for each of the following lags, set a priori between the dose and symptoms series

$$2, 4, 6, 10, 15 \text{ and } 20$$

and the following noise terms, added to the dose series

$$0.25 * N(0, 1), 0.50 * N(0, 1), 1.00 * N(0, 1) \text{ and } 1.50 * N(0, 1).$$

Table 1 shows the percentage of times the tkCCA method correctly identified the a priori set lag for each lag and noise combination. The performance of the tkCCA is perfect for variance (of the noise term) up to 1.0^2, but is slightly less than 100% for variance 1.5^2. However, it should be noted that, for example, with a lag of 20 weeks and a noise term of $1.5 * N(0, 1)$ added to the dose series, the actual Pearson product moment correlation between the symptom series and dose series is

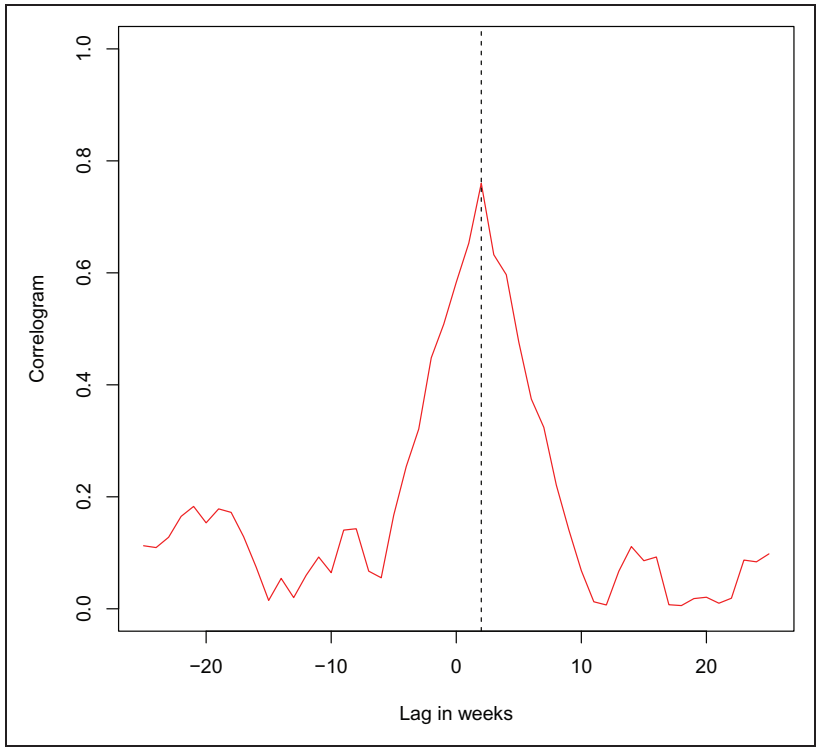

Figure 2. Correlogram from the tkCCA analysis for Figure 1.

Table 1. Performance of the univariate tkCCA in scenarios similar to Figure 1.

Lag → Noise term ↓	2	4	6	10	15	20
0.25*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.50*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
1.00*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
1.50*N(0,1)	98.1%	98.2%	97.8%	98.7%	98.9%	98.1%

approximately 0.01 on the average. Even with such a small correlation between the dose and symptom series, the tkCCA method identifies correctly the a priori set lag for 98.1% of the iterations. Product moment correlation may or may not be a good measure to assess the relationship between the two series, but for comparative purposes, it could be noted that the corresponding correlation with a lag of 2 and noise term of $0.25 * N(0, 1)$ was approximately 0.45. We did not consider a noise term larger than $1.5 * N(0, 1)$ since we think a scenario with Pearson's correlation coefficient between the two series less than 0.01 is not a realistic scenario. In other words, we think that a scenario where there is no relationship between the dose and symptom series should not be used to assess the performance of the tkCCA.

One important feature of the type of dose series presented in Figure 1 and utilized in simulations presented in Table 1 is its highly fluctuating nature. The dose varies almost every time unit. Such high fluctuations in drug intake could be a realistic feature for certain chronic conditions such as HIV infection, but may be not be realistic for certain other chronic conditions such as schizophrenia or Type 2 diabetes. In patients receiving Highly Active Antiretroviral Therapy (HAART) for HIV infection, especially taking a total of 15 or more pills (HIV and non-HIV medications, combined) per day, a highly fluctuating adherence pattern has been observed.[18] Reasons for such high fluctuations in medication adherence could include: being away from home, simply forgetting, falling asleep/sleeping through a dose, feeling depressed/overwhelmed, having problems taking pills at specified times, and/or wanting to avoid side effects.

For other chronic conditions such as Type 2 diabetes or schizophrenia such high fluctuations in dose intake of medications (insulin or an antipsychotic) are not very common. A more common scenario is described as follows: a patient takes the prescribed dose for a few weeks or a few months, sees a substantial improvement in the symptoms, and then drops the dose level to a fraction of the prescribed dose (e.g. half a pill or quarter of a pill) for a few weeks. His/her symptoms may not re-emerge immediately, and so after a few weeks on the reduced dose, the patient may drop it further (e.g. to quarter of a pill or no pills at all). After an extended period of this reduced dose period, the symptoms may re-emerge gradually. The key point is that during the ''back to normal'' period, the patient may consistently take ''half a pill'' for a few weeks and then, for example, drop it down to ''no pills at all'' for a few weeks, before the symptoms slowly re-emerge. In the case of a patient with schizophrenia, for example, the families may intervene when certain symptoms re-appear, and patient may re-start to take at least part of the prescribed dose level, and so on. Thus instead of fluctuating the dose every week, the decrease (''tapering off'') and increase of dose levels would be more gradual. Figure 3 depicts an example of such a scenario for a particular patient.

We used Monte Carlo simulations to assess the performance of the tkCCA for cases similar to that shown in Figure 3. The prescribed initial-response–dose level that we considered for this simulation study was 3 units (mg or g), and the fractional dose levels were

$$0.0, 0.5, 1.0, 1.5, 2.0 \text{ and } 2.5.$$

The time period on each dose level was chosen from

$$0, 2, 3, 4, 5, 6, 7, 8 \text{ and } 10 \text{ weeks}.$$

A dose level was assigned randomly to the time periods 2, 3, 4 and 6, 7, 8 two times, a dose level was assigned randomly assigned a period of 5 weeks six times, and time periods of 0 and 10 week length one time each, so that the total time period would add up to

$$(2 \times 2) + (3 \times 2) + (4 \times 2) + (6 \times 2) + (7 \times 2) + (8 \times 2) + (5 \times 6) + (0 \times 1) + (10 \times 1) = 100 \text{ weeks}.$$

Note that any time period that was repeated was not necessarily on the same dose levels. For example, the patient represented at each iteration will be steadily on a particular dose level for six weeks, twice. However, the dose levels for each of these six-week period need not be the same. Finally, the symptom series was derived by scaling the dose series by a factor of 3, introducing a lag, and adding a noise term. The a priori lags and the noise terms considered for this set of simulations were the same as the ones considered for the previous simulation study (corresponding to Table 1).

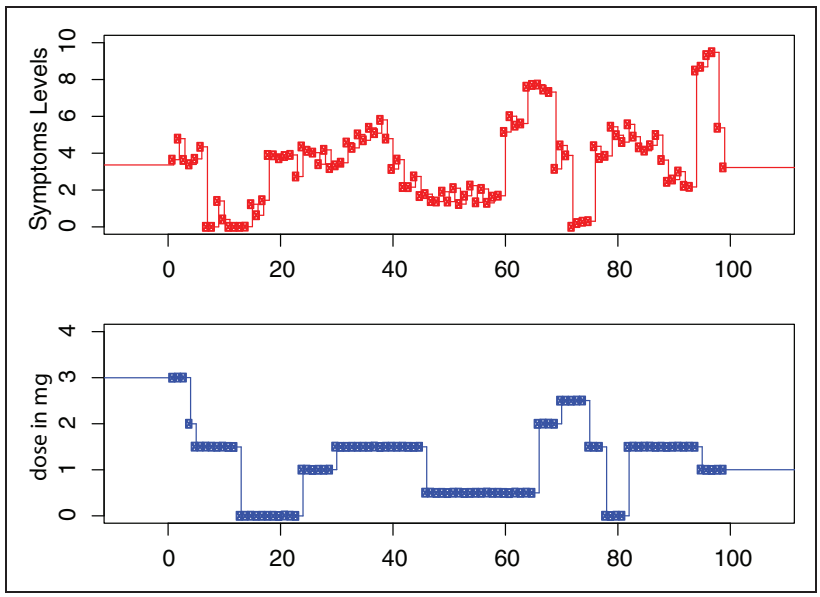

Figure 3. Simulated plot of symptoms and adherence levels for a patient (with a hypothetical brain disorder) with low fluctuations of adherence: Univariate time series example.

Figure 3 shows an example of a pair of the series generated with a lag of two weeks between them and a noise terms of $0.5 * N(0, 1)$. More detailed R codes used to generate the two series in Figure 3 is given below:

```
xx ← sample(c(0, 0.5, 1.0, 1.5, 2.0, 2.5, 3), 20, replace = T) # dose levels
yy ← sample(c(rep(c(4,5,6),2), rep(c(3,5,7),2), rep(c(2,5,8),2), c(0, 10))) #time periods
zz ← 0.5 * rnorm(20)for(i in 1 : length(xx)){
for(i in 1 : length(xx)){
if(i == 1){dd ← rep(xx[1], yy[1]); ee ← rep(zz[1], yy[1])}
if(i > 1){dd ← c(dd, rep(xx[i], yy[i])); ee ← c(ee, rep(zz[i], yy[i]))}}
dd[1 : 4] ← 3
dose ← c(dd[3 : 100], dd[1 : 2]) + 0.0025 * rnorm(100) # setting lag of 2 weeks
symps ← dd * 3 + ee + 0.5 * rnorm(100) # adding the noise term 0.5 * N(0, 1)
```

The results from this simulation study is given in Table 2. In this case, the tkCCA performs perfectly for all the lags and noise terms considered. That is, the method identifies the lag correctly 100% of all the 10, 000 iterations for each lag and noise term. When the lag between the two series is 20 weeks, the performance lowers to approximately 99% for a noise term of $3.5 * N(0, 1)$; such a fluctuation in symptoms is highly unrealistic.

For the above two simulations studies, either the symptoms series were scaled by a factor of 1/3 to get the dose series, or the dose series were scaled by a factor of 3 to get the symptom series. In order to see whether the results depended on scale factors, we ran simulations with scale factors of 1/2 and

Table 2. Performance of the univariate tkCCA in scenarios similar to Figure 3.

Lag → Noise term ↓	2	4	6	10	15	20
0.25*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.50*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
1.00*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
1.50*N(0,1)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

1/4 (or correspondingly 2 and 4, respectively). The results obtained (not shown here) were very similar to those in Tables 1 and 2.

3.2 Multivariate time series example

In this section, we illustrate the tkCCA method for multivariate time series examples. We generated time series for symptoms and adherence levels (dose in mg) for 30 hypothetical patients. Each of the 30 time series were generated by adding Gaussian noise to the univariate time series illustrated in Figure 1 in section 3.1. The lag between the two sets of time series was again 2 weeks. As in the univariate case, the goal of the analysis was to check whether the tkCCA identifies the lag correctly.

The simulated multivariate time series for symptoms and adherence levels are shown in Figure 4, top and bottom panels respectively. As in the univariate case, the lags that we considered for the tkCCA method, equations (7) to (10) are

$$-25,\ -24,\ -23, \ldots, 0, \ldots, 23, 24 \text{ and } 25.$$

The correlogram based on the absolute values of $\rho(\tau)$ in equation (11) is shown in Figure 5. As in the univariate case, the peak of the correlogram is at a lag of two weeks, and so, indeed the tkCCA method detects the lag correctly in the multivariate example also.

As in the univariate case, we conducted Monte Carlo simulations for the multivariate case to check whether the performance of the tkCCA method depended on lag or the noise term. As in the above example, we used a sample size of 30. That is, at each Monte Carlo iteration there were 30 times series of symptom measurements and 30 time series of the dose measurements, generated in a manner similar to the above example. The lags that were set a priori and the noise terms introduced for the dose series were the same as in the simulations done for the univariate case. Additionally, in the multivariate case, we considered different scenarios obtained by varying the proportion of the sample with any particular noise term. So, first we considered scenarios where all the 30 dose series had a fixed variance for the noise term, selected from

$$0.25^2, 0.50^2, 1.00^2 \text{ or } 1.50^2.$$

Then we also considered scenarios where the variance of the noise term was not constant across the 30 samples. For example, we considered a scenario where 50% of the sample had $0.25 * N(0, 1)$ as the noise term, 33% of the sample had $0.50 * N(0, 1)$ as the noise term, and 17% of the sample had $1.00 * N(0, 1)$ as the noise term. All other scenarios that we considered are listed in the first two columns of Table 3. The performance of the tkCCA for various noise scenarios and lags are also

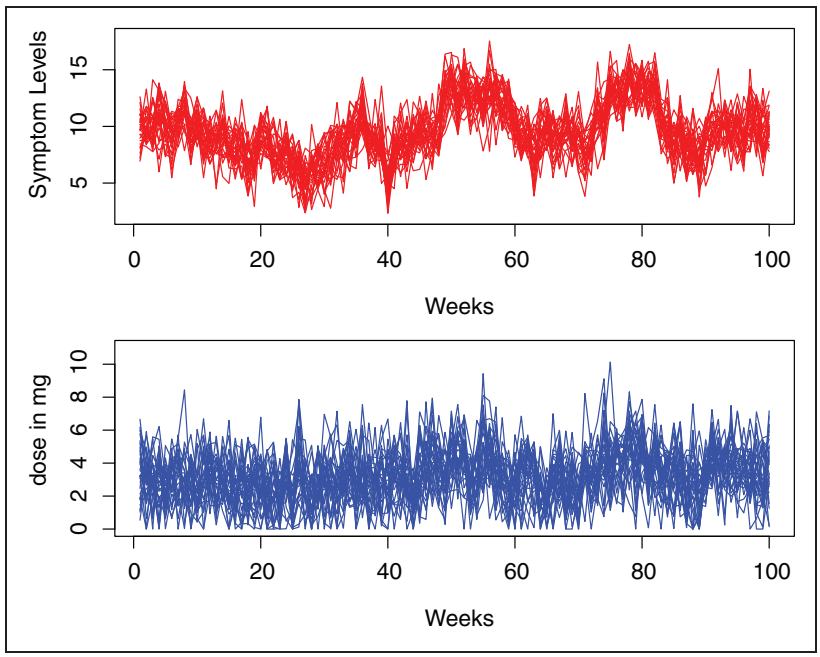

Figure 4. Simulated plot of symptoms and levels for hypothetical patients: Multivariate time series example.

shown in Table 3. When the noise term was kept constant across all the sample, the performance was perfect or near perfect – in all such cases, the a priori set lag was identified correctly by the method 100% (or nearly 100%) of the 10,000 iterations. However, interestingly enough, the performance of the method was far from perfect when we mixed the noise terms among the samples. In such scenarios, this poor performance worsened as the a priori set lag increased. For example, in the scenario where 50%, 33% and 17% of the sample had $0.25 * N(0, 1)$, $0.50 * N(0, 1)$ and $1.00 * N(0, 1)$ as the noise terms, respectively, the performance was 99.7% when the lag was 4, but it decreased substantially to 72.4% when lag was set to 6, and worsened to 49% when the lag was 20.

After finding the poor performance of the multivariate tkCCA in certain scenarios mentioned above and listed in Table 3, we considered an alternate strategy for the multivariate case to see whether it improved the performance. The alternate strategy was to find the lag between the dose and symptom series for each patient separately, considering them as a pair of univariate time series as in the previous section, and then take median of all the 30 lags found as an estimate of the overall lag for the sample. We present the results for this alternate strategy only in the scenarios where the performance of the first strategy was poor. The performance based on this alternate strategy, shown in parenthesis in Table 3, was near perfect (near 100%).

Time series used in all the simulations done in this subsection so far were similar to the ones seen in Figure 1, where the dose series fluctuated very rapidly. Finally, we also evaluated the performance of the tkCCA via simulations in samples (of size 30), where the time series were similar to the ones shown in Figure 3. The lags and noise term scenarios were the same as in Table 3. The performance of the multivariate tkCCA in all scenarios was very good for this set of simulations, as seen in Table 4.

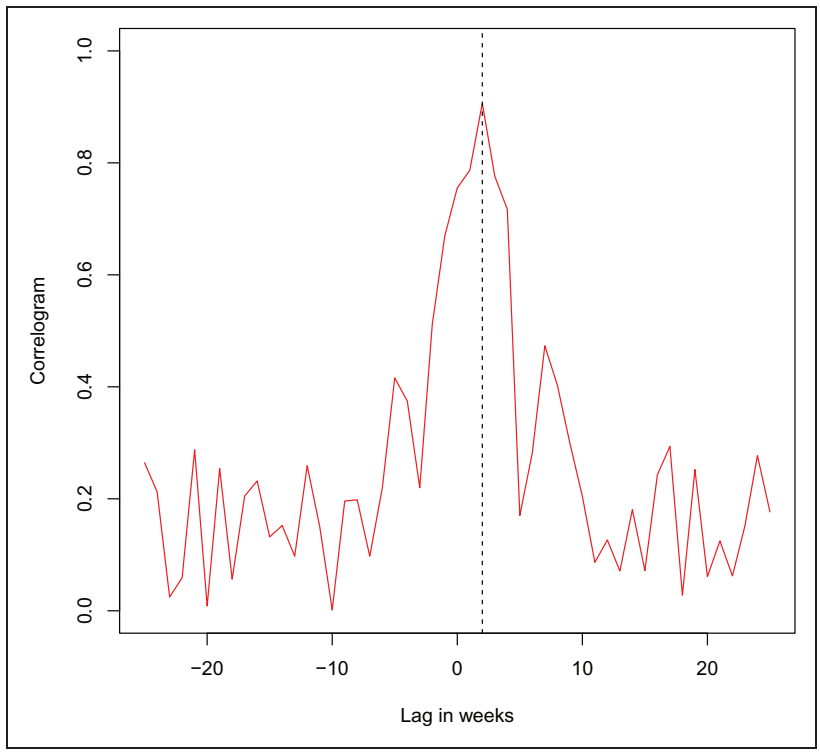

Figure 5. Correlogram from the tkCCA analysis for Figure 4.

The poor performance of the tkCCA seen in some of the scenarios above could be either because those scenarios were not realistic or could be because of inherent limitation of the method itself. However, to be on the safer side, and also since we saw markedly better results by analyzing the sample as separate pairs of univariate time series and taking median of the lags among all patients, we recommend this alternate strategy for clinical applications of the tkCCA.

The complete R codes used for the illustration of the tkCCA method in this section are available upon request from the corresponding author.

4 The tkCCA versus alternate analytic methods for adherence

Having presented the new method, the tkCCA, and evaluated its performance in the previous section, we consider in more detail why the new method would be better than certain other currently existing methods that could be employed to study the effects of non-adherence. Our basic arguments are essentially the same as in MacCallum et al. (2002) paper,[9] but adapted to the examples that we considered in the previous section.

Consider the symptom and dose time series that we looked in subsection 3.1 and plotted in Figure 1. The Pearson product moment correlation, r, between the two time series is 0.37 and the corresponding two-sided p-value based on a t-statistic obtained using the formula

Table 3. Performance of the multivariate tkCCA in scenarios similar to Figure 1.

Noise term ↓	Lag → / Noise proportion ↓	2	4	6	10	15	20
0.25*N(0,1)	100%						
0.50*N(0,1)	0%						
1.00*N(0,1)	0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
1.50*N(0,1)	0%						
0.25*N(0,1)	0%						
0.50*N(0,1)	0%						
1.00*N(0,1)	100%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
1.50*N(0,1)	0%						
0.25*N(0,1)	50%						
0.50*N(0,1)	33%						
1.00*N(0,1)	17%	100.0%	99.7%	72.4%	42.3%	50.7%	48.9%
1.50*N(0,1)	0%			(100.0%)	(99.4%)	(100.0%)	(100.0%)
0.25*N(0,1)	50%						
0.50*N(0,1)	17%						
1.00*N(0,1)	33%	100.0%	99.9%	70.7%	43.4%	51.1%	48.7%
1.50*N(0,1)	0%			(99.6%)	(99.5%)	(99.5%)	(99.5%)
0.25*N(0,1)	50%						
0.50*N(0,1)	0%						
1.00*N(0,1)	33%	99.8%	99.7%	66.0%	41.9%	50.0%	46.5%
1.50*N(0,1)	17%			(99.5%)	(99.2%)	(100.0%)	(100.0%)
0.25*N(0,1)	50%						
0.50*N(0,1)	0%						
1.00*N(0,1)	17%	100.0%	99.8%	67.8%	42.1%	50.0%	45.2%
1.50*N(0,1)	33%			(100.0%)	(99.7%)	(99.6%)	(100.0%)
0.25*N(0,1)	0%						
0.50*N(0,1)	0%						
1.00*N(0,1)	100%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%
1.50*N(0,1)	0%						
0.25*N(0,1)	0%						
0.50*N(0,1)	0%						
1.00*N(0,1)	0%	99.9%	100.0%	99.9%	99.8%	99.4%	99.5%
1.50*N(0,1)	100%						

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \quad n = \text{sample size},$$

is 0.0002 (t-statistic = 3.93). If we now dichotomize the dose series using a median split and use a t-test to compare the mean symptom scores between the two dose categories, we get the t-statistic equal to 1.81 and p-value equal to 0.0737. The correlation after dichotomization using a biserial correlation coefficient was 0.18. Clearly, there is a substantial loss of effect size due to dichotomization, which in turn affected the test of statistical significance. The t-statistic dropped

Table 4. Performance of the multivariate tkCCA in scenarios similar to Figure 3.

Noise term ↓	Lag → / Noise proportion ↓	2	4	6	10	15	20
0.25*N(0,1)	100%						
0.50*N(0,1)	0%						
1.00*N(0,1)	0%	99.7%	99.4%	98.6%	97.5%	97.5%	97.8%
1.50*N(0,1)	0%						
0.25*N(0,1)	0%						
0.50*N(0,1)	100%						
1.00*N(0,1)	0%	99.9%	99.5%	98.9%	98.3%	97.1%	97.7%
1.50*N(0,1)	0%						
0.25*N(0,1)	50%						
0.50*N(0,1)	33%						
1.00*N(0,1)	17%	99.7%	99.3%	98.9%	97.6%	97.6%	97.5%
1.50*N(0,1)	0%						
0.25*N(0,1)	50%						
0.50*N(0,1)	17%						
1.00*N(0,1)	33%	99.7%	99.2%	98.7%	97.4%	97.6%	97.5%
1.50*N(0,1)	0%						
0.25*N(0,1)	50%						
0.50*N(0,1)	0%						
1.00*N(0,1)	33%	99.7%	99.3%	98.7%	97.7%	97.7%	97.6%
1.50*N(0,1)	17%						
0.25*N(0,1)	50%						
0.50*N(0,1)	0%						
1.00*N(0,1)	17%	99.7%	99.2%	98.8%	97.5%	97.5%	97.7%
1.50*N(0,1)	33%						
0.25*N(0,1)	0%						
0.50*N(0,1)	0%						
1.00*N(0,1)	100%	99.8%	99.4%	98.6%	97.7%	96.9%	97.4%
1.50*N(0,1)	0%						
0.25*N(0,1)	0%						
0.50*N(0,1)	0%						
1.00*N(0,1)	0%	99.8%	99.2%	98.2%	97.6%	96.5%	96.9%
1.50*N(0,1)	100%						

from 3.92 prior to dichotomization to 1.81 after, and statistical significance (at 0.05 level) was lost. This loss of statistical significance can be attributed directly to loss of statistical power. To illustrate the loss of power, we generated 10,000 pairs of symptom and dose series as in subsection 3.1 and obtained two p-values at each iteration, one before dichotomization and one after, as in the example aforesaid. Then we estimated the empirical power at 0.05 significance level by calculating the proportion of p-values less than 0.05. The empirical power prior to dichotomization was 99.6% and with a median split dichotomization was 94.1%. For points of dichotomization further away from the median, the empirical power decreased further. For example, when the split point was at

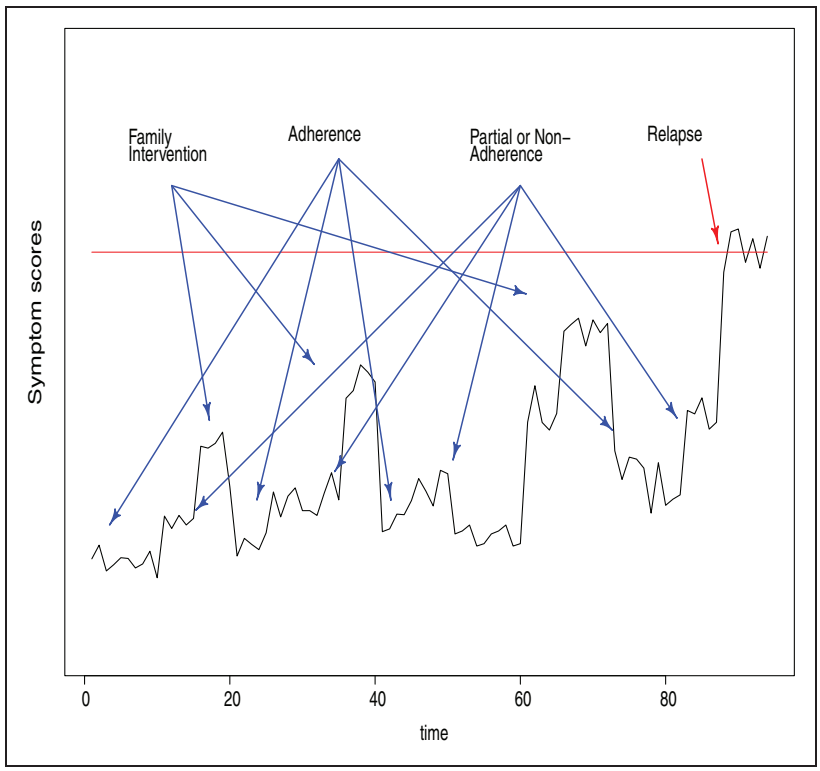

Figure 6. A typical trajectory of symptom progression. Red line represents a hypothetical symptom score at which relapse is considered to occur.

the 90th percentile of the dose series at each iteration, the empirical power was 74.5%. Although the tkCCA and currently used methods are not exactly based on Pearson or biserial correlation, the reduction in the correlation as seen in the above illustration weakens the efficiency of the methods based on dichotomized variables.

Our group[19,20] and others have used survival analysis and Cox proportional hazards regression fruitfully to examine the relationship between non-adherence and relapse. Entering adherence as a time-dependent covariate allows for modeling of variable levels of adherence over time. However, the need to dichotomize outcome (not-relapsed versus relapsed) is a limitation from both a statistical and most importantly, from a clinical perspective. For most illnesses, relapse is not an all-or-none phenomena. Instead, symptom levels increase, sometimes erratically, over time after medication discontinuation. To distinguish between transient symptom exacerbations and a true persistent return of symptoms, most investigators set the level of symptoms needed to meet relapse criteria at a high level. Thus, results for adherence analyses using survival analysis and Cox proportional hazards regression are often clinically informative regarding return of severe, but not intermediate, symptom levels, even though these intermediate levels of symptom return are often clinically important. Figure 6 presents a hypothetical (but realistic) example of the clinical limitations of relying on dichotomous relapse definitions for capturing clinical outcomes. The patient depicted experienced several periods of meaningful symptom exacerbations. Some of these triggered family members to encourage adherence and these efforts resulted in symptom alleviation due to increased

adherence. Eventually, symptom levels reached relapse criteria severity. Conclusions based solely on the exacerbation that met relapse criteria could be misleading as the relationships between non-adherence and earlier symptom exacerbations would not be considered.

5 Performance in the presence of missing data

Missing values, which occur even in well-conducted longitudinal studies, can bias the estimates obtained from such studies. The distribution of the missingness, often called the missingness mechanism,[21,22] is usually classified as three types: Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR). If the missingness does not depend on the outcomes or any covariates (e.g. the patient missed a visit because there was a hailstorm on the day of the appointment), then the mechanism is MCAR. If the missingness depends on the covariates and observed outcomes (e.g. the patient drops out of the study for a few visits after a particular visit when the symptoms were very severe), then the mechanism is MAR. Any violation of MAR falls under the MNAR category. We will be focusing only on MCAR and MAR as there are very few, if any, reasonably good options for dealing with missingness when the mechanism is MNAR. When the mechanism is MCAR, many of the ad hoc methods for imputation such as Last Observation Carried Forward (LOCF) or Mean Imputation (MeI) would work fine. But, in longitudinal clinical studies (especially adherence studies), rarely can the missingness be classified as MCAR. Instead, MAR assumption would be more appropriate for such studies. Under the MAR assumption, ad hoc imputation methods typically give very biased estimates, but an approach based on multiple imputations (MuIs)[22] would provide much better (that is, less biased) estimates. In this section, we explore via simulations the performance of the tkCCA method when missing values are imputed via LOCF, MeI or MuI methods under the MCAR and MAR mechanisms.

All simulations were based on Monte Carlo methods. We used 10,000 Monte Carlo iterations for each simulation. We took the same simulated time series with lag 2 given in section 3.1 and at each Monte Carlo iteration created a new symptoms series and dose series by adding Gaussian white noise with zero mean and standard deviation equalling 0.25. Thus, the time series at each iteration had 100 time points as in the original example.

We imputed missing values assuming either a MCAR mechanism or a MAR mechanism. To simulate missingness under the MCAR mechanism, at each iteration we randomly picked N time points and made the values to be null (that is, missing) at these time points. For our simulation study, we considered $N = 5, 15, 25, 30, 35, 40, 45,$ and 50. To simulate the MAR mechanism, we nullified (that is, set to missing) the value in the time series right after the time point where the value of the symptom series exceeded a particular threhsold value. One of the cut-off values that we considered was 13.5. Thus, in this case, for example, at a particular iteration, if symptom level value was above 13.5 (say 14.0) at the $27th$ time point, then we set the values for the $28th$ time point to be missing. The other cut-off values that we considered were 13.0, 12.5, 12.0, 11.5, 11.0, 10.0 and 9.0. The number of missing values corresponding to the above cut-offs were on the average 4, 9, 14, 16, 19, 22, 32, and 54.

The imputation methods we considered were LOCF, MeI and MuI. LOCF, as the name implies, imputes missing value by carrying forward the last observed non-missing value. There are a few variants of MeI for time series. The variant that we considered was the imputation by the mean of the last observed and the next observed non-missing value. In adherence studies, even if the patient is not completely adherent to the medication, it is still sometimes possible to have the patient make the appointment for the other assessments. Sometimes, even when the patient misses the appointment for a particular visit, by design, there may be other longitudinal variables that may not be missing. Examples of such variables include family reports or even self-reports by patients, which could be

collected over the phone or at a later visit. It is in such cases that the MuIs method (MuI) is particularly useful. MuI works by creating multiple complete data-sets using the joint distribution of all variables in the imputation model. Estimates are obtained for each of these complete data-sets obtained via imputation, and the final estimate is obtained as the average of all the estimates. For our simulations, we created five complete data-sets using MuI. R package *amelia*[23] was used for MuI. For the MuI model, we created two non-missing longitudinal variables *sr* (for self-report) and *fam* for family assessment or family report which were somewhat related to the positive symptoms scores (before assigning missing values) at each Monte Carlo Iteration by using the following code in R:

$$\texttt{sr} \leftarrow \texttt{pos.symps} * (3/2) + 2.5 * \texttt{rnorm}(100)$$
$$\texttt{fam} \leftarrow \texttt{pos.symps} * (5/4) + 3.5 * \texttt{rnorm}(100)$$

Results from the Tables 5 and 6 show that tkCCA performs quite well with the two ad hoc methods, LOCF and MeI, if the missing value mechanism is MCAR. Among the two, MeI has a slight edge over LOCF. Under the MCAR assumption, with LOCF, tkCCA identifies the lag correctly more than 90% of the time even when approximately 35% of the data is missing; with MeI, it identifies the lag accurately more than 90% of the time even when approximately 50% of the data are missing. Hence, based on our simulations, under the MCAR assumption it does not seem to be necessary to use MuIs.

Table 5. Missing value mechanism: MCAR, imputation method: LOCF.

No. of missing values	Percent correct	Mean lag	SD	MSE
$N = 5$	100.00	2.0000	0.0000	0.0000
$N = 15$	99.98	2.0002	0.0141	0.0002
$N = 25$	99.25	2.0075	0.0863	0.0075
$N = 30$	96.70	2.0330	0.1786	0.0330
$N = 35$	92.88	2.0712	0.2572	0.0712
$N = 40$	85.27	2.1473	0.3544	0.1473
$N = 45$	74.82	2.2518	0.4341	0.2518
$N = 50$	60.07	2.3993	0.4898	0.3993

Table 6. Missing value mechanism: MCAR, imputation method: Mean Imputation.

No. of missing values	Percent correct	Mean lag	SD	MSE
$N = 5$	100.00	2.0000	0.0000	0.0000
$N = 15$	100.00	2.0000	0.0000	0.0000
$N = 25$	100.00	2.0000	0.0000	0.0000
$N = 30$	99.95	1.9995	0.0224	0.0005
$N = 35$	99.74	1.9976	0.0509	0.0026
$N = 40$	98.96	1.9916	0.1016	0.0104
$N = 45$	96.72	1.9788	0.1799	0.0328
$N = 50$	93.83	1.9629	0.2456	0.0617

As expected, under the MAR assumption, the performance of the ad hoc methods are as not good as it is under the MCAR assumption (Tables 7–9), although we have to note that they still perform reasonably well when the proportion of missing values is approximately 25% or less. As the proportion of missing values increases beyond 25%, the performance of the LOCF and MeI

Table 7. Missing value mechanism: MAR, imputation method: LOCF.

Cut-off	Mean no. of missing values	Percent correct	Mean lag	SD	MSE
> 13.5	4	100.00	2.0000	0.0000	0.0000
> 13.0	9	99.98	2.0002	0.0141	0.0002
> 12.5	14	99.82	2.0014	0.0548	0.0030
> 12.0	16	98.87	2.0094	0.1154	0.0134
> 11.5	19	95.24	2.0402	0.2396	0.0590
> 11.0	22	91.44	2.0509	0.3174	0.1033
> 10.0	32	88.14	2.0809	0.4445	0.2041
> 9.0	54	27.48	2.7006	0.4841	0.7252

Table 8. Missing value mechanism: MAR, imputation method: mean imputation.

Cut-off	Mean no. of missing values	Percent correct	Mean lag	SD	MSE
>13.5	4	100.00	2.0000	0.0000	0.0000
>13.0	9	99.92	1.9994	0.0283	0.0008
>12.5	14	98.43	1.9827	0.1319	0.0177
>12.0	16	98.23	1.9817	0.1475	0.0221
>11.5	19	97.79	1.9873	0.1246	0.0157
>11.0	22	95.20	2.0056	0.2190	0.0480
>10.0	32	84.47	2.1389	0.3688	0.1553
>9.0	54	27.48	2.7006	0.4841	0.7252

Table 9. Missing value mechanism: MAR, imputation method: multiple imputation.

Cut-off	Mean no. of missing values	Percent correct	Mean lag	SD	MSE
>13.5	4	100.00	2.0000	0.0000	0.0000
>13.0	9	99.92	1.9998	0.0282	0.0008
>12.5	14	99.77	1.9999	0.0480	0.0023
>12.0	16	99.53	2.0005	0.0686	0.0047
>11.5	19	98.24	1.9931	0.1380	0.0191
>11.0	22	95.76	1.9746	0.2130	0.0460
>10.0	32	94.65	1.9624	0.3191	0.1032
>9.0	54	84.15	1.7848	0.6823	0.5118

methods worsens quite rapidly. Based on our simulations, MuIs is clearly the method of choice for the tkCCA if the assumption is MAR and the proportion of missing values is greater than 25%.

6 Real data example

In this section, we illustrate the application of the tkCCA to a real data example. Medication non-adherence (either total refusal of medication or taking only part of prescribed medication) is common in all branches of medicine among patients requiring prolonged maintenance treatment.[24–26] Our example presents data on the effects of medication non-adherence by patients with a first episode of schizophrenia-spectrum disorder. We focused upon the effects of non-adherence on the severity of psychotic symptoms (also known as positive symptoms), one of the core symptoms of schizophrenia.[27,28]

Data are from one treatment arm (the antipsychotic risperidone prescribed using a flexible dosing schedule) of a longitudinal study of first episode schizophrenia-spectrum disorders funded by the National Institute of Mental Health, USA. Twenty-seven subjects who responded to their randomly assigned risperidone medication within 16 weeks of starting initial treatment were eligible to be continued on their successful medication for a total treatment period of 3 years. Follow-up data were censored at the time that subjects left their randomly assigned treatment for any reason. Subjects were encouraged to remain on the risperidone dose associated with their acute response to treatment during the maintenance treatment phase of the study. The data were collected on actual (as opposed to prescribed) doses of medication taken. The total amount of risperidone taken during a period of interest was divided by the length of the period to obtain an average daily dose taken for that period. For this analysis, we examined psychotic symptom levels and risperidone dose taken using a monthly time frame. Because we wished to examine maintenance treatment issues, we set the baseline for the analyses as the study week at which each subject met a priori criteria for response from their initial psychotic episode. We imputed for missing values in each of the individual time series by interpolating the values at the neighboring time points. The time series for psychotic symptoms and risperidone dose are shown in Figure 7. We analyzed the sample as separate pairs of univariate time series to estimate lag for each patient using the univariate tkCCA method and took the median of these lags to be the estimate of the lag for the overall sample. The lags that we considered for this analysis were 0, 4, 8, 12 up to 100 weeks.

The lag estimate for the overall sample was obtained as 24 weeks. This suggests that, on the average, psychotic symptoms would reappear around 6 months as the average medication dose taken decreases due to partial or full non-adherence from the dose that produced response of the symptoms of the initial illness episode. Leave-one-out cross validation (LOOCV) method, by leaving out the data from one patient at a time, was employed to estimate the precision of the above lag estimate. LOOCV gave a standard deviation of 6.9 for the lag estimate. Also, 91% of the iterations in the LOOCV method showed the lag estimate to be 24.

7 Conclusion

We presented the temporal kernel canonical correlation analysis (tkCCA) method developed by Bieβmann and colleagues and showed how it can be utilized to answer certain interesting questions in adherence studies. The tkCCA is an extension of the Canonical Correlation Analysis for a set of time series with non-instantaneous coupling. We used simulations to generate univariate and multivariate time series corresponding to symptoms and adherence levels in patients with a hypothetical brain disorder, and illustrated how the tkCCA correctly finds the lag between the

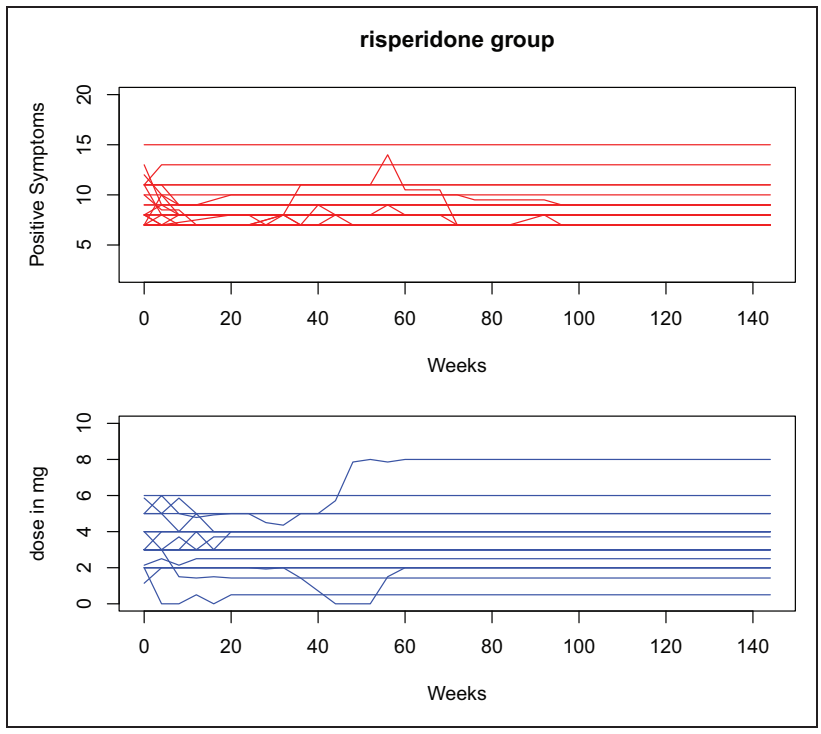

Figure 7. Plot of positive symptoms (that is, psychotic symptoms) and adherence levels for patients with first episode of a schizophrenia-spectrum disorder treated with risperidone: Real data example.

two sets of times series. Thus, for example, the tkCCA may be used as an analytical tool to predict the number of weeks that takes for the symptoms of a patient to relapse when his or her adherence falls off from the optimal response dose. We also examined the performance of the tkCCA under the MCAR and MAR missing value mechanisms with imputations methods such as LOCF, MeI, and MuIs. The tkCCA with LOCF or MeIs performed quite well under MCAR assumption but performed poorly under the MAR assumption. Nevertheless, under the MAR assumption, the tkCCA with MuIs performed reasonably well. Further, we applied the tkCCA to a real data example of psychotic symptoms and dose levels obtained from a study based on subjects with a first episode of schizophrenia, schizophreniform or schizoaffective disorder on risperidone treatment, and showed that for the subjects in this study, on the average, the re-appearance of the psychotic symptoms as the adherence levels falls off was at about 6 months.

One limitation of the tkCCA is the computational time taken for running the method since it involves inversion of a large matrix as part of a generalized eigenvalue problem. However, computational burden is not forbiddingly high. For any iteration in our simulations, it never took more than 2 minutes to run the tkCCA on a standard PC. As processor speeds increase in the future, this would be less of a limitation.

Another limitation of the tkCCA method is that it assumes a constant lag between the two time series considered. In medicine, disease progression may change the time lag between events. For example, relapse off treatment may occur faster as patients have an illness longer. The tkCCA method in its current format is not capable of dealing with such situations. Extension of the tkCCA method to allow time varying lags will be an area of future research.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Majnu John's research was supported by NIH grants P30MH090590 and P50MH080173.

References

1. Osterberg L and Blaschke T. Adherence to medication. *N Engl J Med* 2005; **353**: 487–497.
2. McDonnell PJ and Jacobs MR. Hospital admissions resulting from preventable adverse drug reactions. *Ann Pharmacother* 2002; **36**: 1331–1336.
3. Schiff GD, Fung S, Speroff T, et al. Decompensated heart failure: symptoms, patterns of onset, and contributing factors. *Am J Med* 2003; **114**: 625–630.
4. Senst BL, Achusim LE, Genest RP, et al. Practical approach to determining costs and frequency of adverse drug events in a health care network. *Am J Health Syst Pharm* 2001; **58**: 1126–1132.
5. Misdrahi D, Llorca PM, Lancon C, et al. Compliance in schizophrenia: predictive factors, therapeutical considerations and research implications. *Encephale* 2002; **28**: 266–272. (In French).
6. Rodgers PT and Ruffin DM. Medication nonadherence: Part II a pilot study in patients with congestive heart failure. *Manage Care Interface* 1998; **11**: 67–69, 75.
7. Ventura J, Lukoff D, Nuechterlein KH, et al. Brief Psychiatric Rating Scale (BPRS), expanded version: scales, anchor points, and administration manual. *Int J Meth Psychiatr Res* 1993; **3**: 227243.
8. Nuechterlein KH, Miklowitz DJ, Ventura J, et al. Classifying episodes in schizophrenia and bipolar disorder: criteria for relapse and remission applied to recent onset samples. *Psychiatry Res* 2006; **144**: 153166.
9. MacCallum RC, Zhang S, Preacher KJ, et al. On the practice of dichotomization of quantitative variables. *Psychol Meth* 2002; **7**: 19–40.
10. Johnson RA and Wichern DW. *Applied multivariate statistical analysis*, 6th edn. Pearson: Prentice Hall, 2007.
11. Akaho S. A kernel method for canonical correlational analysis. In: *Proceedings of the international meeting of the Psychometric Society (IMPS)*. Osaka, Japan, Springer-Verlag, 15–19 July 2001.
12. Bach FR and Jordan MI. Kernel independent component analysis. *J Mach Learn Res* 2002; **3**: 1–48.
13. Fukumizu K, Bach FR and Gretton A. Statistical consistency of kernel CCA. *J Mach Learn Res* 2007; **8**: 361–383.
14. Bießmann F, Meinecke FC, Gretton A, et al. Temporal kernel CCA and its application in multimodal neuronal data analysis. *Mach Learn* May 2010; **79**(1–2): 5–27.
15. Murayama Y, Bießmann F, Meinecke FC, et al. Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA. *Magn Reson Imag* 2010; **28**: 1095–1103.
16. Hasselman B. geigen: Calculate Generalized Eigenvalues of a Matrix Pair. R-package version 1.3, http://CRAN.R-project.org/package=geigen (2015, accessed May 2015).
17. Siegel SJ. Extended release drug delivery strategies in psychiatry: theory to practice. *Psychiatry (Edgmont)* 2005; **2**: 22–31.
18. Skrajner MJ, Camp CJ, Heckman TG, et al. A Fluctuating pattern of over- and under-adherence for haart: a case study from a videophone intervention project. *Clin Gerontol*. Vol. 34: 2011, pp.144–153.
19. Robinson D, Woerner M, Alvir J, Ma J, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Archive General Psychiatr* 1999; **56**: 241–247.
20. Robinson DG, Woerner MG, Alvir JM., et al. Predictors of medication discontinuation by patients with first-episode schizophrenia and schizoaffective disorder. *Schizophrenia Res* 2002; **57**: 209219.
21. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**: 581–592.
22. Schafer JL and Graham JW. Missing data: our view of the state of the art. *Psychol Meth* 2002; **7**: 147–177.
23. Honaker J, Joseph A, King G, et al. 1998-2002. AMELIA: A program for missing data, http://gking.harvard.edu/amelia (accessed May 2015).
24. Horne R, Weinman J, Barber N, et al. Concordance, adherence and compliance in medicine taking: a conceptual map and research priorities. London: National Coordinating Centre for NHS Service Delivery and Organisation (NCCSDO), http://www.sdo.lshtm.ac.uk/files (2006, accessed May 2015).
25. Horne R. Adherence to medication: a review of existing research. In: Myers L and Midence K (eds) *Adherence to treatment in medical conditions*. London: Harwood Academic, 1998, pp.285–310.
26. Velligan DI, Weiden PJ, Sajatovic M, et al. Expert consensus panel on adherence problems in serious and persistent mental illness. The expert consensus guideline series: adherence problems in patients with serious and persistent mental illness. *J Clin Psychiatr* 2009; **70**: 1–46.
27. Gaebel W, Riesbeck M, Wlwer W, et al. German Study Group on First-Episode Schizophrenia. Relapse prevention in first-episode schizophrenia-maintenance vs intermittent drug treatment with prodrome-based early intervention: results of a randomized controlled trial within the German Research Network on Schizophrenia. *J Clin Psychiatr* 2011; **72**: 205–218.
28. Subotnik KL, Nuechterlein KH, Ventura J, et al. Risperidone nonadherence and return of positive symptoms in the early course of schizophrenia. *Am J Psychiatry* 2011; **168**: 286–292.