IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **38** (2017) 1456–1471 https://doi.org/10.1088/1361-6579/aa724d

Continuous assessment of schizophrenia using heart rate and accelerometer data

Erik Reinertsen[1], Maxim Osipov[2], Chengyu Liu[3], John M Kane[4], Georgios Petrides[4] and Gari D Clifford[1,3]

[1] Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America
[2] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom
[3] Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, United States of America
[4] Department of Psychiatry, The Zucker Hillside Hospital and Hofstra Northwell School of Medicine, Hempstead, NY, United States of America

E-mail: er@gatech.edu

Received 16 February 2017, revised 12 April 2017
Accepted for publication 10 May 2017
Published 27 June 2017

Abstract

Objective: Schizophrenia has been associated with changes in heart rate (HR) and physical activity measures. However, the relationship between analysis window length and classifier accuracy using these features has yet to be quantified. *Approach*: Here we used objective HR and activity data to classify contiguous days of data as belonging to a schizophrenia patient or a healthy control. HR and physical activity recordings were made on 12 medicated subjects with schizophrenia and 12 healthy controls. Features derived from these data included classical statistical characteristics, rest-activity metrics, transfer entropy, and multiscale fuzzy entropy. We varied the analysis window length from two to eight days, and selected features via minimal-redundancy-maximal-relevance. A support vector machine was trained to classify schizophrenia from control windows on a daily basis. Model performance was assessed via subject-wise leave-one-out-crossfold-validation. *Main results*: An analysis window length of eight days resulted in an area under a receiver operating characteristic curve (AUC) of 0.96. Reducing the analysis window length to two days only lowered the AUC to 0.91. The type of most predictive features varied with analysis window length. *Significance*: Our results suggest continuous tracking of subjects with schizophrenia over short time scales may be sufficient to estimate illness severity on a daily basis.

Keywords: schizophrenia, machine learning, support vector machine, mhealth, mobile health, heart rate, entropy

(Some figures may appear in colour only in the online journal)

1. Introduction

Schizophrenia is a chronic psychiatric disease and global health problem with a lifetime prevalence of 4.0/1000 (Saha *et al* 2005). It is among the most disabling and economically catastrophic disorders; the overall cost of schizophrenia to the US in 2002 was estimated at 62.7B due to clinical care, medication, and unemployment (Wu *et al* 2005). Onset usually occurs in early adult years. Schizophrenia is characterized by delusions, hallucinations, disorganization of speech and behavior, and a higher rate of co-occurring psychiatric disorders. Depression prevalence is 25%, which is higher than the rate in the general population (Buckley *et al* 2009), and lifetime risk of suicide is 5% (Hor and Taylor 2010, Kahn *et al* 2015).

Schizophrenia is diagnosed via clinical interview, in which the psychiatrist asks the patient to self-report characteristic symptoms, and assesses if social and/or occupational dysfunction has occurred for at least six months. Unfortunately schizophrenia impairs insight which hinders the accuracy of self-reporting. Even if a patient is diagnosed and treated, medication adherence is poor and seriously consequential (Byerly *et al* 2007).

Differences in power spectral and/or entropy measures of both activity and heart rate (HR) have been reported in schizophrenia patients versus healthy controls (Bär *et al* 2008, Hauge *et al* 2011, Wulff *et al* 2012). These sophisticated measures include frequency components and information theory complexity, contain more information than classical statistical characteristics such as the mean of a signal, and reflect changes in the autonomic nervous system (ANS).

Previously we reported differences in ANS function, sleep patterns, and locomotor activity in subjects with schizophrenia versus healthy controls. We accurately distinguished schizophrenia patients from controls by training a machine learning algorithm with features derived from HR and locomotor activity selected from the highest-quality ten days of data recorded from a body worn patch (Osipov *et al* 2015). Using both HR and activity features improved classification accuracy compared to using features from just one data type. However, the optimal timescale of data measurement necessary to accurately estimate the severity of schizophrenic symptoms remains unknown. Understanding this could contribute to the passive, objective, and near real-time monitoring of schizophrenia patients to detect early signs of illness relapse, medication adherence, or treatment efficacy.

Here we vary the analysis window length of recorded HR and locomotor activity data, extract features from these data, train a support vector machine to distinguish if a patient has a diagnosis of schizophrenia or is a healthy control, and evaluate classifier performance for differing analysis window lengths.

2. Methods

2.1. Participants and data collection

16 outpatient subjects diagnosed with schizophrenia but in symptomatic remission, and 19 healthy control volunteers without a history of mental disorders were recruited for the study. All subjects were unemployed. Although the prevalence of schizophrenia is only 4/1000 in the general population, this balanced cohort was appropriate for our study aim: to develop a method for estimating illness severity in subjects with an established diagnosis of schizophrenia. Age and gender did not significantly differ among the two groups, as assessed via a two-sided Student's *t*-test and Fisher's exact test, respectively. Subjects diagnosed with

schizophrenia were taking anti-psychotic medications including Olanzapine, Risperidone, Aripiprazole, Perphenazine, Fluphenazine, Ziprasidone, Haloperidol, and Quetiapine.

HR and locomotor activity were monitored for 3–4 weeks using a disposable adhesive patch sensor worn on the chest and manufactured by Proteus Biomedical (Redwood City, CA). Electrocardiogram (ECG)-derived HR data were collected every 10 min by calculating mean HR over 15 s intervals. Accelerometry-derived locomotor activity data were collected every 5 min by calculating mean acceleration over 15 s intervals. Data were transmitted to a mobile phone via Bluetooth and further uploaded to a central server for processing.

2.2. Data pre-processing

Matlab R2016a (Mathworks, Natick, MA) was used to analyze HR (beats per minute, or BPM) and locomotor activity (normalized units $\in [0, 1]$) time series data, which were collected with variable recording rates and lengths. Data collected at an insufficient sampling rate—after an interval exceeding $1.5\times$ the sampling period, which equals 15 min for HR and 7.5 min for activity—were discarded. Additionally, HR values lower than 20 BPM or higher than 160 BPM were labeled as low-quality and removed. A day was considered to have sufficient data if it contained both (a) at least 50 HR data points and (b) at least 50 locomotor activity data points. Subjects with fewer than ten days of sufficient data were removed.

Four subjects with schizophrenia lacked sufficient amounts of data, and twelve subjects with schizophrenia possessed sufficient amounts of data. Prior to analysis, excess control subjects were removed at random to ensure both groups had 12 patients (24 subjects total). HR data were re-sampled to 10-min intervals, and activity data were re-sampled to 5-min intervals, both via linear interpolation.

Data pre-processing, feature extraction, classifier training, and model evaluation was performed for sliding contiguous windows of length of w days where $w \in [2, 4, 6, 8]$. Days with at least 50 HR and at least 50 activity data were considered viable. Days with fewer data were skipped. The ith analysis window started at day d_i and ended at day $d_i + w - 1$ where d_i is the ith day of data. i started at 1 and incremented to the last day of viable data for a subject. If a day was not viable, i.e. lacked sufficient data, no features were extracted, schizophrenia status was not estimated, and the algorithm incremented to the next day.

2.3. Statistical characteristics

HR and locomotor activity are affected in schizophrenia so we calculated classical statistical characteristics of HR and activity, including mean, median, mode, standard deviation (σ) and interquartile range (IQR) (Bär *et al* 2008, Hauge *et al* 2011, Rachow *et al* 2011).

2.4. Rest-activity characteristics

Circadian rhythm disruption has been shown to play a significant role in schizophrenia (Wulff *et al* 2012). We calculated rest-activity characteristics of HR and activity, including mean level during the least active five hours (L5), mean level during the most active ten hours (M10), relative amplitude (RA, equation (1)), interday stability (IS, equation (2)) and intraday variability (IV, equation (3)) (Witting *et al* 1990, Van Someren *et al* 1999):

$$RA = \frac{M10 - L5}{M10 + L5} \tag{1}$$

$$IS = \frac{n \sum_{h=1}^{p} (\bar{x_h} - \bar{x})^2}{p \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2}$$

$$IV = \frac{n \sum_{i=2}^{n} (x_i - x_{i-1})^2}{(n-1) \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3}$$

where n is the total number of data, p is the number of data per day, x_h are hourly means, x is the mean of all data, and x_i represents individual data points.

2.5. Behavioral features

The same features were extracted as described in our previous work, including basic statistical characteristics, rest-activity characteristics, and multiscale transfer entropy (MTE) between HR and activity data. 36 total features were calculated, which are described in Osipov *et al* (2015). However, we calculated multiscale fuzzy entropy, a novel metric of sequence complexity, instead of multiscale entropy (MSE).

2.5.1. Multiscale fuzzy entropy. Activity disorganization is one of the diagnostic criteria for schizophrenia (Kahn *et al* 2015). Hauge *et al* (2011) indicated that entropy can be used to estimate such disorganization. Moreover, our group has demonstrated that MSE applied to actigraphy provided information on disorganization in schizophrenic patients (Osipov *et al* 2015). MSE is the calculation of sample entropy (SampEn) H for a range of time scales of original signal (Richman and Moorman 2000, Costa *et al* 2002). This metric of signal complexity is derived from the negative logarithm of the conditional probability of the appearance of longer patterns in a signal, considering the presence of a shorter pattern:

$$H(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)} \tag{4}$$

where m is the template length, r is the radius of similarity or distance threshold between patterns, $A^m(r)$ is a probability of matching a template of length $m + 1$, $B^m(r)$ is the probability of matching a template of length m, and N is the number of elements in the time series. Two patterns of length m are considered similar if each point of a pattern in one part of the signal is within a normalized distance r from the respective point in the other part of the signal.

To vary time scales of data used to calculate entropy, the original signal was coarse-grained to a lower temporal sampling frequency (i.e. by half for each coarse-graining). Coarse-grained time series were constructed by averaging the data points within non-overlapping windows of increasing length τ. For the τth time scale, each element of the coarse-grained time series, $y_j^{(\tau)}$, is given by:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \tag{5}$$

where τ represents the scale factor and $1 \leqslant j \leqslant N/\tau$. The length of each coarse-grained time series is N/τ. The first time scale corresponds to the 0th coarse graining and is just the original time series.

Although MSE can differ by illness status, SampEn can change significantly and/or non-monotonically with small changes in parameter values and thus exhibits poor statistical stability. Additionally, SampEn only accounts for similar patterns with similar amplitudes, not similar patterns with different amplitudes. These shortcomings are addressed by the recently proposed fuzzy entropy, $\mathcal{H}_{\text{fuzzy}}$ (Lee *et al* 2001, Liu *et al* 2013, Li *et al* 2013, 2014).

To calculate $\mathcal{H}_{\text{fuzzy}}$, the binary Heaviside classifier in SampEn is replaced with a continuous membership degree between 0 and 1, based on Zadeh's concepts of fuzzy set theory (Chen *et al* 2007). $\mathcal{H}_{\text{fuzzy}}$ accounts for both local and global sequence regularity and is more robust to noise than SampEn. Multiscale fuzzy entropy (MFE) is determined by calculating $\mathcal{H}_{\text{fuzzy}}$ for coarse-graining a time series several times, in the same fashion as MSE described above.

Prior work indicates that lower scales of MSE provide the best features for training a classifier to distinguish schizophrenic from healthy subjects (Osipov *et al* 2015). Therefore we evaluated MSE of HR and activity data for the first four time scales (in the coarse-graining sense, not window length). Furthermore, a classifier trained on MFE of heart rate data from patients with heart failure outperformed a classifier trained on MSE of the same data (Liu *et al* 2013). Thus, we calculated MFE of HR and activity data for the first four time scales, using parameter values $m_{\text{HR}} = 3$, $m_{\text{act}} = 2$, $r_{\text{HR}} = 0.1$, and $r_{\text{act}} = 0.15$.

2.5.2. Feature selection. In order to reduce the number of features used and minimize overfitting, a feature selection approach was necessary. Features were z-scored, discretized, and ranked via minimum redundancy maximum relevance (mRMR) criteria (Peng *et al* 2005). This approach simultaneously minimizes mutual information between individual features x in a feature set S:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \tag{6}$$

and maximizes mutual information between features x and classes (sometimes referred to as labels or targets) c:

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{7}$$

where $I(x;y)$ is the mutual information between variables x and y, and $p(x)$, $p(y)$, and $p(x, y)$ are probability densities of these variables:

$$I(x; y) = \int p(x, y) \log \frac{p(x, y)}{p(x)\, p(y)} \mathrm{d}x\mathrm{d}y. \tag{8}$$

These two constraints for D and R are combined into one expression $\phi(D, R)$ which is optimized as follows:

$$\max \phi(D, R) = D - R. \tag{9}$$

The mRMR algorithm ordered features by values of ϕ from highest to lowest. Hereafter, 'most predictive' refers to the subset of features with the highest values of ϕ.

Statistical characteristics, rest-activity characteristics, MTE, and MFE were calculated and 36 features total were ranked via mRMR. The most predictive i features (where $i \in 1, 2, ..., m$) were used to train a machine learning algorithm.

2.6. Classification of schizophrenia status among subjects

Subjects were classified as either having a diagnosis of schizophrenia or being healthy, using `libsvm`, an open-source support vector machine (SVM) library (Chang and Lin 2011). The two–dimensional matrix of features consisted of W windows by m features, and the one-dimensional array of labels consisted of W binary labels. Each analysis window was labeled as 1 if belonging to a schizophrenia patient, or 0 if belonging to a healthy control. A Gaussian radial basis function kernel with $\gamma = 0.0312$ was selected based on previous work (Osipov *et al* 2015).

The most predictive features identified via mRMR were used to train the SVM, which output probability estimates of a subject being labeled as schizophrenic, $P(SZ)$, rather than healthy. Classifier performance was assessed via subject-wise leave-one-out crossfold validation (LOOCV; given N patients, $N-1$ patients are used to train the classifier and the remaining patient is used as the test set), and various attributes of classifier performance was calculated, including area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. The number of most predictive features (number of features maximizing the AUC) was also determined.

Pre-processing, feature extraction, and classifier assessment was performed for analysis window lengths of two, four, six, and eight days. The SVM was trained using windows, not on individual days or subjects.

3. Results

The number of features resulting in the maximum AUC, and the maximum AUC value itself, both differed with analysis window length (figure 1). With an analysis window length of two days, the model achieved a maximum AUC of 0.91 using five most predictive features. With an analysis window length of eight days, the model achieved a maximum AUC of 0.96 using 20 most predictive features. Table 1 lists other classifier performance metrics for varying window lengths.

Box plots of the most predictive features (i.e. the combination of features which maximized the training AUC) for two-day and eight-day analysis windows are shown in figures 2 and 3 respectively. For two-day analysis windows, the three most predictive features in order of more to less predictive are (1) the standard deviation of activity (σ_{act}), 2) the multiscale fuzzy entropy of heart rate at the first time scale ($\text{MFE}_{\text{HR},1}$), and 3) the mode of activity (Mo_{act}).

For eight-day analysis windows the eleven most predictive features in order of more to less predictive are 1) $\text{MFE}_{\text{HR},1}$, 2) IQR_{act}, 3) Mo_{act}, 4) the multiscale transfer entropy from activity to HR at the first time scale (coarse-graining) ($\text{MTE}_{\text{act}\rightarrow\text{HR},1}$), 5) σ_{act}, 6) $\text{MFE}_{\text{act},4}$, 7) $\text{MFE}_{\text{HR},4}$, 8) the intraday variability of activity (IV_{act}), 9) $\text{MTE}_{\text{act}\rightarrow\text{HR},2}$, 10) $\text{MFE}_{\text{act},2}$, and 11) the relative amplitude of activity (RA_{act}).

Probability density estimates of classifier output (estimated probability of a window of data belonging to a subject with schizophrenia) for schizophrenia patients distinctly differed from those of control subjects. This difference was large for both two-day (figure 4(a)) and eight-day analysis windows (figure 4(b)).

ROC curves showed a positive correlation between analysis window length and AUC (figure 5). AUC values ranged from 0.91 for two-day windows to 0.96 for eight-day windows. Increasing window length increased most metrics of classifier performance (table 1) with only a few exceptions. Increasing window length from two to four days reduced sensitivity from 0.89 to 0.85 and NPV from 0.79 to 0.76. Increasing analysis window length from four to six

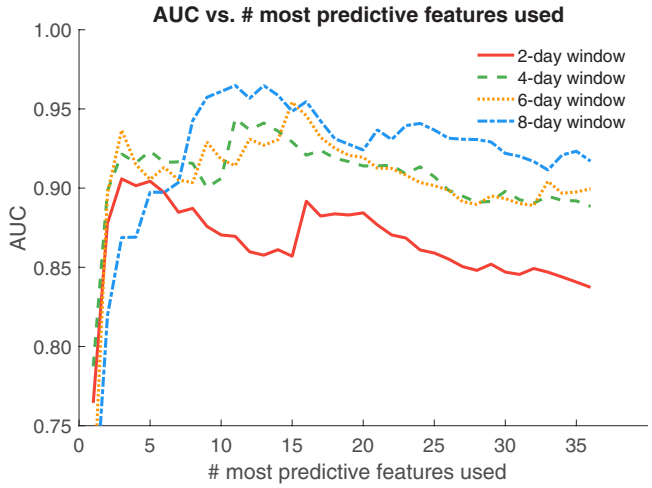

Figure 1. AUC versus number of most predictive features, selected out of 36 total features via mRMR, used to train the SVM. The blue line represents two-day analysis windows and the red line represents eight-day analysis windows. The maximum AUC for two-day analysis windows is 0.92 using the five most predictive features, and the maximum AUC for eight-day analysis windows is 0.96 using the 20 most predictive features.

Table 1. Classifier performance metrics versus window length, using both HR and activity features. PPV indicates positive predictive value and NPV indicates negative predictive value.

	Window length (days)			
Metric	2	4	6	8
AUC	0.91	0.94	0.95	0.96
Accuracy	0.85	0.89	0.89	0.91
Sensitivity	0.89	0.85	0.87	0.87
Specificity	0.76	0.98	0.94	0.99
PPV	0.87	0.98	0.97	0.99
NPV	0.79	0.76	0.77	0.76

days reduced specificity from 0.98 to 0.94, and PPV from 0.98 to 0.97. Increasing window length from six to eight days reduced NPV from 0.77 to 0.76.

To assess the relative contribution of each type of feature (i.e. heart rate, locomotor activity, or both combined) we compared classifier AUC versus feature type for analysis window lengths of two or eight days. With an analysis window length of two days, features from HR resulted in a classifier AUC of 0.80, features from locomotor activity resulted in a classifier AUC of 0.85, and features from both HR and locomotor activity resulted in a classifier AUC of 0.92 (table 2). These results demonstrate an improvement classifier performance when using features from both HR and locomotor activity data, compared to using features from either category alone.

HR and activity data, estimated $P(SZ)$, optimal classifier thresholds, and data quantity were visualized against 24 h intervals into the study for a representative subject with schizophrenia (figure 6(a)) and a representative healthy control subject (figure 6(b)).

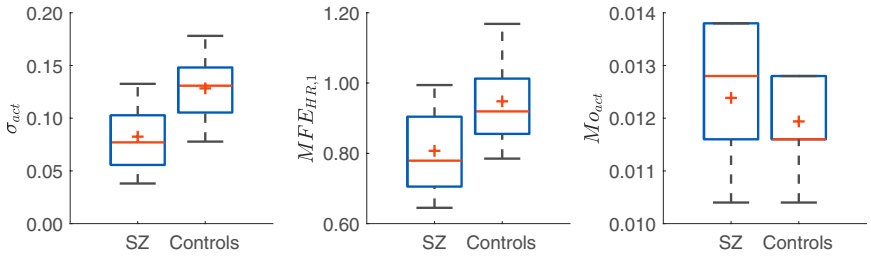

Figure 2. Box plots of most predictive features selected via mRMR using two-day analysis windows. The SZ label on the *x*-axis indicates features from schizophrenia patients. These three features in combination maximized the training AUC. The red + indicates the mean, the middle horizontal red line indicates the median, the blue box denotes the interquartile range (IQR) flanked by the 25th and 75th percentiles, and the vertical lines outside of the box indicate the 9th and 91th percentiles. The median value of every feature significantly differed by schizophrenia status, with $P < 0.05$ calculated via two-sided Wilcoxon rank-sum test.

4. Discussion

We built on previous work using features derived from HR and activity to classify medicated schizophrenia patients from healthy controls. Here we evaluated the relationship between analysis window length and classifier accuracy.

Using two-day analysis windows, a maximum AUC of 0.91 was achieved using the three most predictive features. Using eight-day analysis windows, a maximum AUC of 0.96 was achieved using the 11 most predictive features (figure 1). Both AUCs sharply increased as initial features were added. For the two-day model, adding more than three features steadily lowered the AUC, aside from a slight rise in classifier performance when moving from 16 to 17 features. On the other hand, for 4-, 6-, and 8-day models the AUC stayed close to 0.90 for a wide range of 10 to 25 features. This may be due to the fact that, for the two-day model, the number of data points is low enough that higher-dimensional models begin to fit more noise.

The ranking of features via mRMR criteria depended on analysis window length (figures 2 and 3). The most predictive feature from two-day analysis windows was σ_{act} with a median $\pm$ IQR of 0.08 ± 0.05 units for the schizophrenic group compared to 0.13 ± 0.04 units for the control group. This difference was statistically significant ($P < 0.05$, Wilcoxon rank-sum test), contrasting with results from Hauge *et al* (2011), which also reported significantly lower levels of mean activity in schizophrenic patients. Our algorithm did not identify mean activity as a predictive feature. However, that work reported the evaluation of long-term patients from an open ward, whereas our study evaluated outpatients in relative symptomatic remission. Illness severity and behavioral markers may have differed.

The 2th most predictive feature for two-day analysis windows was $\text{MFE}_{\text{HR},1}$, with a median $\pm$ IQR of 0.95 ± 0.50 in schizophrenic patients and 1.30 ± 0.40 in controls. The 3th most predictive feature for two-day analysis windows was Mo_{act}, with a median $\pm$ IQR of 0.013 ± 0.002 in schizophrenic patients and 0.012 ± 0.001 in controls. All three most predictive features for two-day analysis windows differed significantly by schizophrenia status ($P < 0.05$, Wilcoxon rank-sum test).

The most predictive feature for eight-day windows was $\text{MFE}_{\text{HR},1}$, with a median $\pm$ IQR of 0.89 ± 0.46 for schizophrenic patients and 1.24 ± 0.19 for controls (3). This difference was consistent and statistically significant for both two- and eight-day analysis windows ($P < 0.05$, Wilcoxon rank-sum test). Other predictive features for eight-day analysis windows

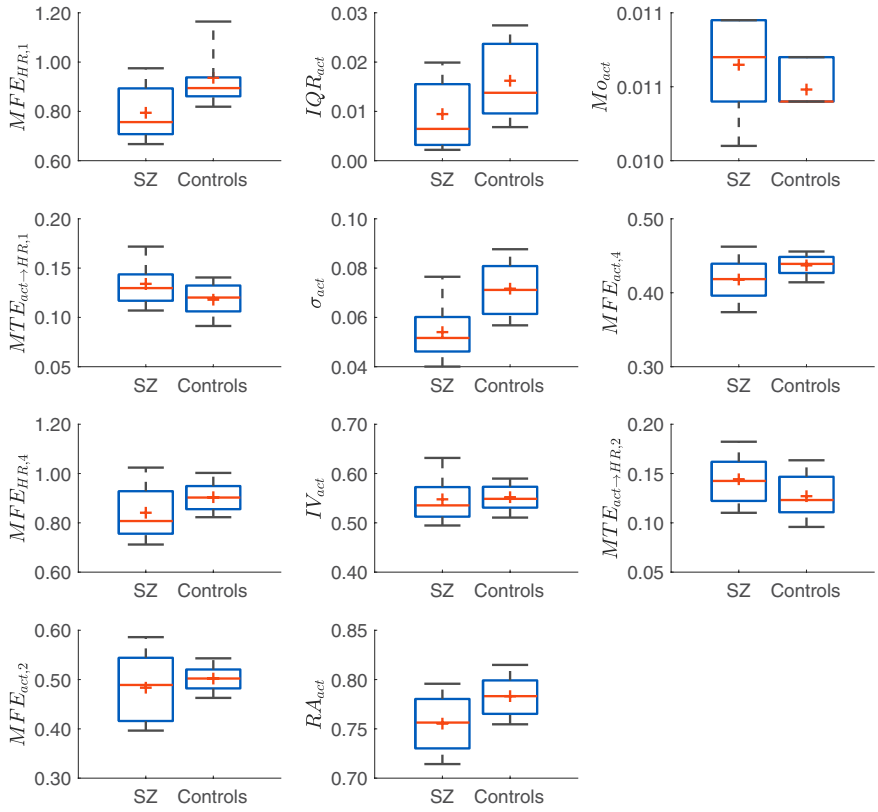

Figure 3. Box plots of most predictive features selected via mRMR using eight-day analysis windows. The SZ label on the x-axis indicates features from schizophrenia patients. These eleven features in combination maximized the training AUC. The red $+$ indicates the mean, the middle horizontal red line indicates the median, the blue box denotes the interquartile range (IQR) flanked by the 25th and 75th percentiles, and the vertical lines outside of the box indicate the 9th and 91th percentiles. The median value of every feature significantly differed by schizophrenia status, with $P < 0.05$ calculated via two-sided Wilcoxon rank-sum test.

represent the spread of the distribution of activity, such as $\mathrm{IQR}_{\mathrm{act}}$ and σ_{act}. We note skewness, kurtosis, mean, or median of activity were not predictive features. Furthermore, aside from $\mathrm{MFE}_{\mathrm{HR},1}$, no other HR statistics were predictive.

$\mathrm{MTE}_{\mathrm{act}\rightarrow\mathrm{HR},i}$ denotes the transfer entropy, or amount of uncertainty reduced in future values of HR by knowing the past values of activity given past values of HR, at a time scale i. For two-day analysis windows, $\mathrm{MTE}_{\mathrm{act}\rightarrow\mathrm{HR},i}$ was not a predictive feature for any time scale. However, when using eight-day analysis windows, $\mathrm{MTE}_{\mathrm{act}\rightarrow\mathrm{HR},i}$ was a predictive feature for $i = 1$ and $i = 2$. When using two-day analysis windows, $\mathrm{MTE}_{\mathrm{act}\rightarrow\mathrm{HR},i}$ is still predictive; however, it is not as predictive as the most predictive three features described above. These results suggest activity-driven changes in HR become more predictive at longer time scales.

We used the mRMR framework to select features on the basis of maximizing predictivity (relevance), and minimizing collinearity (redundancy) (Peng *et al* 2005). Although our most-predictive features were all separable in a univariate sense, features should be selected by predictivity rather than significance for classification tasks (Lo *et al* 2015). Features may be inseparable in a univariate sense but strongly separable in higher dimensions. On the other

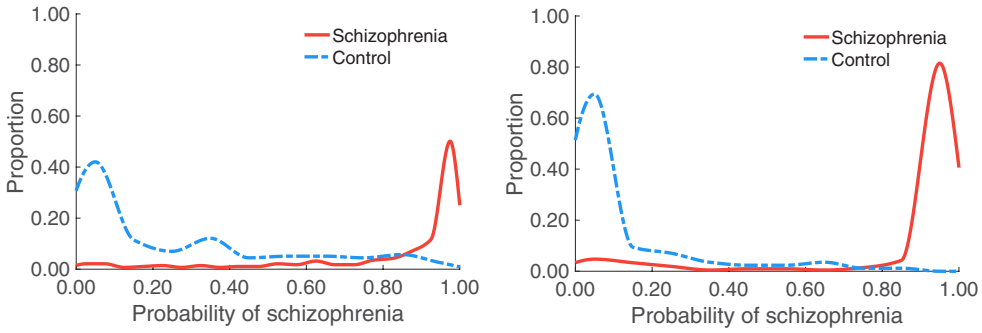

Figure 4. Probability density estimates of classifier output (estimated probability of a window of data belonging to a subject with schizophrenia) using (a) two-day and (b) eight-day analysis windows. Classifier output is on the *x*-axis, and proportion is on the *y*-axis; all *y*-values for a class sum to unity. Output from schizophrenic subjects is in red and control subjects in in blue.

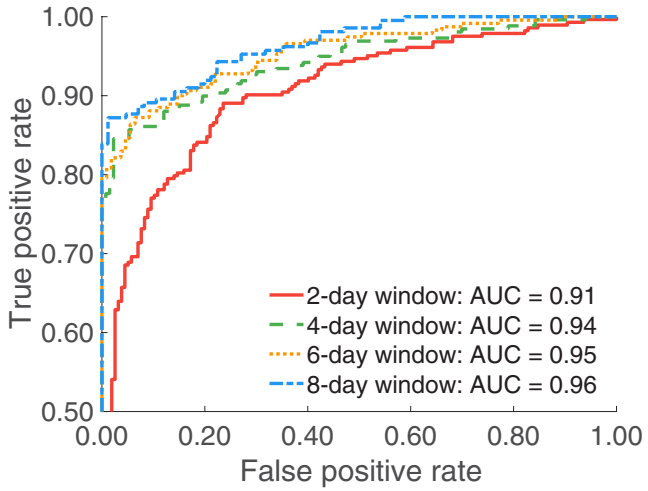

Figure 5. Receiver operating characteristic (ROC) curves vary with analysis window length. Blue, red, yellow and purple denote two, four, six, and eight-day windows respectively. The *y*-axis is the true positive rate, or sensitivity. The *x*-axis is the false positive rate, or 1 − specificity.

Table 2. Area under the ROC curve (AUC) versus window length and feature type used to train support vector machine. AUCs calculated via leave-one-out-cross-validation (LOOCV).

	Window length (days)	
Feature type(s)	2	8
Heart rate (HR)	0.84	0.90
Activity	0.86	0.89
HR and activity	0.91	0.96

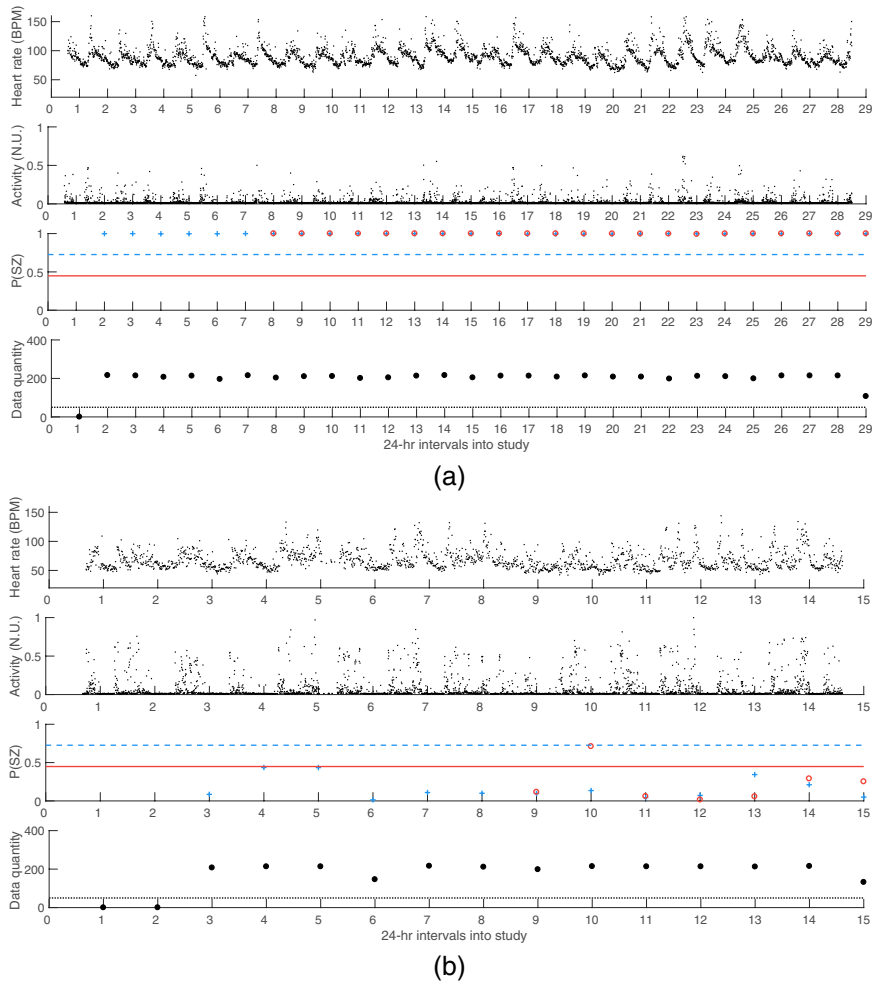

Figure 6. HR data (top row), activity data (second row), classifier output (probability of schizophrenia, or $P(SZ)$; third row), and data quantity versus time (bottom row) for a (a) schizophrenia patient and a (b) healthy control subject. Heart rate is in beats per minute (BPM), activity is in normalized units (N.U.), $P(SZ)$ is a probability, and data quantity is in raw counts. $P(SZ)$ for a window length of two days is shown by the red $+$'s. The classifier threshold for a two-day window is $P(SZ) = 0.45$, is shown by the red solid line. $P(SZ)$ for a window length of eight days is shown by the blue circles. The classifier threshold for a window length of eight days, $P(SZ) = 0.73$, is shown by the blue dashed line. On the data quantity plot, the minimum data quantity (at least 50 HR and at least 50 activity data) required to make a estimate of $P(SZ)$ on a given day is shown by the black dotted line.

hand, a feature may significantly differ between classes in a univariate or even multivariate sense, but may still not contribute predictive accuracy to a model.

With two-day analysis windows, the SVM classifier produced grossly distinct distributions of output for schizophrenia patients versus healthy controls, demonstrating excellent classifier performance (figure 4(a)). However, a few analysis windows were mislabeled. Using eight-day analysis windows resulted in similar excellent classifier performance evidenced by large

separability of classifier output between windows of data from schizophrenia patients and healthy controls (figure 4(b)).

AUC increased with analysis window length, from 0.91 for two days, up to 0.96 for eight days (figure 5). In our previous work, an SVM was trained using the ten days missing the least data from each subject, and achieved an AUC of 0.99 (Osipov *et al* 2015). In this work, classifier performance was slightly lower because we did not pick 'best' days among several weeks of data. Rather, we slid an analysis window through all days from all patients and trained a classifier on individual windows. This approach better represents a realistic scenario in which only a few days of patient data are obtained.

We also assessed how analysis window length affects the sensitivity, specificity, positive predictive value, negative predict value, and accuracy of the classifier for different window lengths (table 1). Specificity exceeded 0.90 for four-, six-, and eight-day windows. The AUC uniformly increased with window length. Compared to two-day windows, four-day windows resulted in a higher AUC, accuracy, and specificity, but sensitivity dropped from 0.89 to 0.85. A decrease in sensitivity is undesirable when the cost of missing a true positive is high (e.g. suicide). A monitoring system to track illness severity of patients already diagnosed with schizophrenia might thus prioritize sensitivity, although doing so increases false positives.

Analysis window length determined which features are most predictive, mediates classifier performance, perhaps due to differing time scale of relevant physiology and behavior. Prior work on the association between schizophrenia and disturbances in the ANS as measured by heart rate variability metrics has shown a similar dependence on physiologically and behaviorally relevant time scales (Bär *et al* 2008, Rachow *et al* 2011). A feature calculated from two days of data may contain information about systems mediated by circadian rhythms and sleep. Alternatively, a feature calculated from eight days of data may contain information about social activity, behavior, and lower-frequency physiological dynamics. On the scale of days to weeks, social drivers of behavior (i.e. cadence of the work week) that are not present on the scale of hours may become more apparent in these data. Schizophrenia patients typically have disrupted social routines (Kahn *et al* 2015). Recordings on the order of months may be necessary to measure physiology or behavior mediated by hormonal cycles or seasons. The selection of analysis window length thus may be an important consideration in the design of studies for monitoring patients with mental illness.

We assessed the relative contribution of HR and locomotor activity features to classifier accuracy. Previously we have shown a combination of HR and activity features outperforms either HR or activity features alone (Osipov *et al* 2015). Here we show the same trend for both two- and eight-day analysis windows (table 2). For two-day analysis windows, using activity features (AUC of 0.86) resulted in better classifier performance than HR features (AUC of 0.84). For eight-day analysis windows, using HR features (AUC of 0.90) resulted in slightly better classifier performance than activity features (AUC of 0.89). Using both HR and activity features together improved the AUC, for both two- and eight-day analysis windows. Locomotor activity and HR are correlated—the former introduces artifact and random error to HR, and HR tends to rise during locomotor activity—yet contribute complementary and non-redundant information about subject behavior that improves the predictive accuracy of a classifier.

Individual data for patients, estimated probabilities of schizophrenia $P(SZ)$, and data quantity for each 24 h interval were visualized for a representative schizophrenia patient (figure 6(a)) and healthy control subject (figure 6(b)). Time series data were re-sampled during signal processing, so overall data quantity per 24 h interval was almost equal between schizophrenia and control subjects. Upon visual inspection, HR appears more periodic for the schizophrenia patient compared to the control. This observation is consistent with $\mathrm{MFE}_{\mathrm{HR},1}$

being significantly lower in schizophrenia patients than in controls for both two- (figure 2) and eight-day analysis windows (figure 3), and with previous reports of lower HR complexity in these patients (Rachow *et al* 2011).

For some schizophrenia patients, estimated $P(SZ)$ occasionally fell below the classification threshold (figure 6(a)). Likewise, for some control subjects, estimated $P(SZ)$ rose above the threshold (figure 6(b)). Due to the lack of more richly labeled data, it is unclear if this misclassification indicates control subjects have schizophrenic-like days and schizophrenia patients have normal-like days. Regardless, averaging strategies could theoretically reduce false positives. The optimal trade-off between sensitivity or specificity is determined by the use case and cost of false positives or false negatives; here we simply maximized the AUC.

We note several limitations of our study. Our sample size was relatively small and limited to patients from one geographic region and institution. Also, several factors such as mental and behavioral state, social status, and medication usage affect HR and activity.

Although we controlled for employment status as a proxy for social routine—which relates to activity and restfulness to some extent—we did not explore other potential confounders that could affect HRV or locomotor activity. Literature suggests potential confounders such as weight, BMI, diet, smoking status, renal function, etc have a moderate effect on HRV. However, the dominant factor by an order of magnitude is the mental response (Bernardi *et al* 2000). The second-largest dominant factor is physical movement (Knoepfli-Lenzin *et al* 2010). In this work we also analyzed locomotor activity, which we have previously shown contains predictive information for classifying subjects by schizophrenia status (Osipov *et al* 2015).

Stress relates to mental state with physiological and behavioral manifestations, and could thus influence our features. Skin conductance, a biomarker for stress mediated by the sympathetic nervous system, has been shown to differ in schizophrenia patients (Bär *et al* 2008). Additionally, cortisol secretion and stress sensitivity may be associated with schizophrenia, or subsequent development of schizophrenia following the prodromal phase of the illness (Holtzman *et al* 2013, Walker *et al* 2013). Stress affects activity as well as other aspects of mobile device usage; Sano and Picard (2013) reported using using screen on, mobility, call or activity level information to distinguish stressed from non-stressed individuals with an accuracy of 75%.

Aside from matching employment status, we did not directly control for stress in our study. Doing so with conscious patients is challenging especially in an ambulatory setting. Additionally, stress invariably accompanies social risk factors such as physical abuse, sexual abuse, maltreatment and bullying, which are associated with increased risk of later schizophrenia (Stilo and Murray 2010).

Evaluating data solely from periods of lower activity or stress based on physiology could reduce confounding from these random variables. Recently we attempted to reduce noise in data and improved classifier performance by selecting quiescent periods of data with lowest median HR, which is a proxy criterion for restfulness (Reinertsen *et al* 2017). More sophisticated change point detection approaches could potentially sort data into parametrically similar segment, and even further increase the signal-to-noise ratio of features derived from data within each segment (Adams and MacKay 2007).

Antipsychotic medications have been reported to exacerbate ANS dysfunction in schizophrenia patients, potentially putting patients at greater risk of cardiac mortality (Rechlin *et al* 1994, Birkhofer *et al* 2013, Huang *et al* 2013). On the other hand, Mondelli *et al* demonstrated that antipsychotic medication can reduce cortisol secretion and normalize HPA-axis hyperactivity in patients suffering from psychosis (Mondelli *et al* 2010). Bär *et al* (2008) did not find significant changes in ANS function after antipsychotics were administered to patients, while

(Henry *et al* 2010) *et al* found that risperidone, valproate, or mood stabilizers did not significantly affect HRV in bipolar or schizophrenia patients. We lacked more detailed information about the type, dose, or adherence of antipsychotic medications taken by patients in our study. We thus cannot claim our results generalize to non-medicated subjects with schizophrenia. However, a classifier affected by a patient's medication could potentially be used to monitor adherence and treatment efficacy.

Illness severity in schizophrenia fluctuates from day to day (Kahn *et al* 2015) and strongly correlates with measures of ANS dysfunction such as HR variability (Bär *et al* 2005, Bär *et al* 2008, Henry *et al* 2010, Montaquila *et al* 2015). Our classifier output varied from day to day (figure 4), was based on measures of ANS dysfunction and behavior, and may have reflected fluctuations in illness severity. However, we lacked detailed information about symptoms, e.g. Brief Psychiatric Rating Scale survey data, which would be necessary for training a classifier to estimate illness severity.

Lastly, some hyper-parameters of our model, such as entropy template length and minimum amount of data per day, were selected from prior work instead of optimized. Classifier performance could likely be improved using techniques such as Bayesian optimization (Ghassemi *et al* 2014, Shahriari *et al* 2016).

5. Conclusion

We evaluated the relationship between analysis window length of data recordings and classifier accuracy in a small cohort of schizophrenia patients and healthy controls. A support vector machine was trained on HR and locomotor activity data obtained via body-worn patches and windowed over varying lengths ranging from two to eight days. A novel metric—multiscale fuzzy entropy—contributed to the predictive accuracy of our model. Our approach accurately classified schizophrenia status in a small cohort of subjects with an AUC of 0.91 for an analysis window length as short as two days, and an AUC of 0.96 for an analysis window length of 8 d. Features selected as most predictive also varied with analysis window length. Our classifier output may have reflected illness severity, although verifying this in future work will require gathering information about symptoms on a daily basis. This work serves as technical proof of the feasibility of HR and activity-based classification of illness severity in the context of schizophrenia.

Acknowledgments and funding

The authors thank Proteus Digital Health for assistance with data collection. This research is funded by Emory University, a Research Council UK Grant EP/G036861/1 (CDT in Healthcare Innovation), a Wellcome Trust Centre Grant No. 098461/Z/12/Z (Sleep, Circadian Rhythms and Neuroscience Institute), and EPSRC grant EP/K020161/1 (Multiscale markers of circadian rhythm changes for monitoring of mental health).

References

Adams R P and MacKay D J C 2007 Bayesian online changepoint detection (arXiv: 0710.3742v1) p 7

Bär K J *et al* 2005 Loss of efferent vagal activity in acute schizophrenia *J. Psychiatric Res.* **39** 519–27

Bär K J *et al* 2008 Relationship between cardiovagal modulation and psychotic state in patients with paranoid schizophrenia *Psychiatry Res.* **157** 255–7

Bernardi L *et al* 2000 Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability *J. Am. Coll. Cardiol.* **35** 1462–9
Birkhofer A *et al* 2013 The deceleration capacity—a new measure of heart rate variability evaluated in patients with schizophrenia and antipsychotic treatment *Eur. Psychiatry* **28** 81–6
Buckley P F *et al* 2009 Psychiatric comorbidities and schizophrenia *Schizophrenia Bull.* **35** 383–402
Byerly M J, Nakonezny P A and Lescouflair E 2007 Antipsychotic medication adherence in schizophrenia *Psychiatric Clinics North Am.* **30** 437–52
Chang C C and Lin C J 2011 LIBSVM: a library for support vector machines *ACM Trans. Intell. Syst. Technol.* **2** 27
Chen W *et al* 2007 Characterization of surface EMG signal based on fuzzy entropy *IEEE Trans. Neural Syst. Rehabil. Eng.* **15** 266–72
Costa M, Goldberger A and Peng C K 2002 Multiscale entropy analysis of complex physiologic time series *Phys. Rev. Lett.* **89** 068102
Ghassemi M *et al* 2014 Global optimization approaches for parameter tuning in biomedical signal processing: a focus of multi-scale entropy *Comput. Cardiol.* **41** 993–6
Hauge E R *et al* 2011 Nonlinear analysis of motor activity shows differences between schizophrenia and depression: a study using Fourier analysis and sample entropy *PLoS One* **6** e16291
Henry B L *et al* 2010 Heart rate variability in bipolar mania and schizophrenia *J. Psychiatric Res.* **44** 168–76
Holtzman C *et al* 2013 Stress and neurodevelopmental processes in the emergence of psychosis *Neuroscience* **249** 172–91
Hor K and Taylor M 2010 Suicide and schizophrenia: a systematic review of rates and risk factors *J. Psychopharmacol.* **20** 81–90
Huang W L *et al* 2013 Impact of antipsychotics and anticholinergics on autonomic modulation in patients with schizophrenia *J. Clin. Psychopharmacol.* **33** 170–7
Kahn R S *et al* 2015 Schizophrenia *Nat. Rev. Dis. Primers* **1** 15067
Knoepfli-Lenzin C *et al* 2010 Effects of a 12-week intervention period with football and running for habitually active men with mild hypertension *Scand. J. Med. Sci. Sports* **20** 72–9
Lee H M *et al* 2001 An efficient fuzzy classifier with feature selection based on fuzzy entropy *IEEE Trans. Syst. Man Cybern.* B **31** 426–32
Li P *et al* 2014 Coupling between short-term heart rate and diastolic period is reduced in heart failure patients as indicated by multivariate entropy analysis *Comput. Cardiol.* **41** 97–100
Li P *et al* 2013 Multiscale multivariate fuzzy entropy analysis *Acta Phys. Sin.* **62** 120512
Liu C *et al* 2013 Analysis of heart rate variability using fuzzy measure entropy *Comput. Biol. Med.* **43** 100–8
Lo A *et al* 2015 Why significant variables aren't automatically good predictors *Proc. Natl Acad. Sci.* 112 13892–7
Mondelli V *et al* 2010 Abnormal cortisol levels during the day and cortisol awakening response in first-episode psychosis: the role of stress and of antipsychotic treatment *Schizophrenia Res.* **116** 234–42
Montaquila J M, Trachik B J and Bedwell J S 2015 Heart rate variability and vagal tone in schizophrenia: a review *J. Psychiatric Res.* **69** 57–66
Osipov M *et al* 2015 Objective identification and analysis of physiological and behavioral signs of schizophrenia *J. Mental Health* **24** 276–82
Peng H, Long F and Ding C 2005 Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226–38
Rachow T *et al* 2011 Nonlinear relationship between electrodermal activity and heart rate variability in patients with acute schizophrenia *Psychophysiology* **48** 1323–32
Rechlin T *et al* 1994 Heart rate variability in schizophrenic patients and changes of autonomic heart rate parameters during treatment with clozapine *Biol. Psychiatry* **35** 888–92
Reinertsen E *et al* 2017 Heart rate-based window segmentation improves accuracy of classifying posttraumatic stress disorder using heart rate variability measures *Physiol. Meas.* **38** 1061
Richman J S and Moorman J R 2000 Physiological time-series analysis using approximate entropy and sample entropy *Am. J. Physiol. Heart Circ. Physiol.* **278** H2039–49
Saha S *et al* 2005 A systematic review of the prevalence of schizophrenia *PLoS Med.* **2** 141
Sano A and Picard R W 2013 Stress recognition using wearable sensors and mobile phones *Humaine Association Conf. on Affective Computing and Intelligent Interaction Stress*

Shahriari B *et al* 2016 Taking the human out of the loop: a review of Bayesian optimization *Proc. IEEE* **104** 148–75

Stilo S A and Murray R M 2010 The epidemiology of schizophrenia: replacing dogma with knowledge *Dialogues Clin. Neurosci.* **12** 305–15

Van Someren E J W *et al* 1999 Bright light therapy: improved sensitivity to its effects on rest-activity rhythms in Alzheimer patients by application of nonparametric methods *J. Biol. Med. Rhythm Res.* **16** 505–18

Walker E F *et al* 2013 Cortisol levels and risk for psychosis: initial findings from the North American prodrome longitudinal study *Biol. Psychiatry* **74** 410–7

Witting W *et al* 1990 Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease *Biol. Psychiatry* **27** 563–72

Wu E Q *et al* 2005 The economic burden of schizophrenia in the United States in 2002 *J. Clin. Psychiatry* **66** 1122–9

Wulff K *et al* 2012 Sleep and circadian rhythm disruption in schizophrenia *Br. J. Psychiatry* **200** 308–16