

# 1 Background

We use digital technologies to monitor continuously over time patients behavior.

1) We would like to leverage this information to study the causal effect of mobile communication (say daily counts of outgoing calls) on mental health scores (say anxiety). The goal of the project is to evaluate using simulations the performance of hidden markov models for zero-inflated count data in estimating the causal effect in N-of-1 time series data.

Mobile health Apps become more and more important in psychiatry. Among mental health patients, there is a great interest in monitoring symptoms with mobile apps [17]. The digital technologies are used to monitor patients behavior over time through collection of the long-term objective data continuously and automatically. These data vary from activity data such as GPS, distance traveled to social communication data such as text messages (eg, the number of outgoing SMS) and phone calls (eg, the number of incoming and outgoing calls, duration outgoing calls). Indeed, using both self-reported information collected by the self-rating questionnaires and the smartphone sensor data offers a unique opportunity to describe the person in terms of his or her lifestyle and behavior at the physical, cognitive, and environmental level. [Cornet and Holden \[2018\]](#), [Torous et al. \[2018\]](#)

Researchers hope to gain insight in the user's clinical outcome through analysis of the variety of personal data. More studies are needed to establish the relationship between mobile measurements and clinical symptoms. This mobile phone-based approach may provide an estimation of physiological and mental state. In addition, it may be valuable to predict the changes in clinical states and investigate causal inferences about state changes in patients (eg, those with affective disorders) [Dogan et al. \[2017\]](#)

Multiple statistical methods were developed for the analysis of time series data of mental health. However, when the observations are a sequence of unbounded counts, using regular time series models that are established on continuous outcomes with continuous distribution are not suitable.<sup>11,12</sup>

The Poisson distribution is a natural choice to describe such data. The most important exclusivity of this distribution is the equality of the mean and the variance. But often this exclusivity will not hold and variance is bigger than the mean, which is called over dispersion. Using mixture models is one way to handle this problem.<sup>12</sup> As when the data are a series of unbounded counts there may be over dispersion, and serial dependency, using independent Poisson random variables are also not appropriate and some models that incorporate this dependency are needed.<sup>12</sup> Hidden Markov model can accommodate both over dispersion and serial dependency by allowing the probability distribution of each observation to depend on

The rapid growth of smart-sensor integration in mobile phones and wearable devices has opened the prospect of increasing access to evidence-based mental health care. In this review, the term sensor-based data includes the quantitative information supplied by the mobile phone and its embedded sensors. Information may range from acceleration to temperature and from light to pressure, but also from number of exchanged short message service (SMS) text messages to number of incoming and outgoing calls. Even if the evidence of association between sensor-based data and psychiatric disorder status and/or severity of psychiatric symptoms is limited and scattered [15-17], it is expected that appropriate management of

these data may initiate a new trend in health care provision characterized by tailored and timely interventions [18].

## 2 Objective

The primary goal of this project is to build a hidden markov models for zero-inflated count data and self-reported categorical data to predict the clinical state of next time point through estimation of the transition probability of clinical states.

Next, we need to evaluate using simulations the performance of different hidden markov models, such as basic HMM, HMM dealing with missing data, HMM for non-stationary time series and other time series model such as GLM, ARMA.

Furthermore, the final goal is to obtain the causal effect between mobile communication time series data and the clinical outcomes.

## 3 Method

## References

- V. P. Cornet and R. J. Holden. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of biomedical informatics*, 77:120–132, 2018.
- E. Dogan, C. Sander, X. Wagner, U. Hegerl, and E. Kohls. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? systematic review. *Journal of medical Internet research*, 19(7):e262, 2017.
- J. Torous, P. Staples, I. Barnett, L. R. Sandoval, M. Keshavan, and J.-P. Onnela. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *NPJ digital medicine*, 1(1):1–9, 2018.