Statistics in Biosciences (2019) 11:207–209
https://doi.org/10.1007/s12561-019-09247-1

GUEST EDITORIAL

Editorial for the Special Issue "*Medical Device Data: Challenges, Statistical Methods and Applications*"

Jaroslaw Harezlak[1] · **Chongzhi Di**[2]

Published online: 24 June 2019

Medical devices have been widely used in biomedical studies for measuring clinical modalities, physiological conditions, environmental exposures, and behavior and lifestyle factors. Their use has exploded in the recent years, partly due to rapid technological advances in areas including wearable and implantable sensors, cellular or wireless-enabled devices and smartphone-based apps, making it feasible and inexpensive to adopt them in large-scale studies. These studies now generate high-resolution data with massive size and complex structures, bringing unprecedented challenges to statistical analysis.

This special issue is dedicated to analysis of medical device data. We are very excited to include twelve interesting papers, representing a wide range of topics in terms of types of devices, scientific context and relevant statistical methodologies. The types of devices include wearable accelerometers and heart rate monitors for measuring physical activity, which is arguably one of the earliest and most widely adopted wearable devices in biomedical research (with seven papers devoted to this topic), in-shoe sensors, electronic medication adherence monitoring devices and smartphone apps. The scientific contexts vary from physical activity epidemiology, aging and cardiovascular diseases to HIV prevention research. A wide variety of topics on relevant statistical methods were covered, including feature extraction, functional data analysis, longitudinal data analysis, clustering and machine learning.

That being said, we recognize that the field of medical devices is extremely broad and evolves rapidly with new technologies. While it is impossible to comprehensively cover every topic, this special issue represents the first attempt to systematically group the statistical community's recent efforts in this diverse area, with

✉ Jaroslaw Harezlak
harezlak@iu.edu

Chongzhi Di
cdi@fredhutch.org

1 Department of Epidemiology and Biostatistics, School of Public Health-Bloomington, Indiana University, 1025 E 7th St., C107, Bloomington, IN 47405, USA

2 Biostatistics Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B500, Seattle, WA 98109, USA

a focus on wearable and mobile devices. Our purpose is to provide a systematic review, foster collaborations and inspire many more statisticians and other quantitative researchers to work in this area.

Karas et al. provide a comprehensive review on challenges and opportunities for analyzing accelerometry data in health research. They discuss problems related to the collection and analysis of raw accelerometry data, including issues arising from data volume and structure, heterogeneity, device calibration, batch effects, data labeling and multiple monitor synchronization. Published solutions are referred to if available, and their points are illustrated using the Developmental Epidemiological Cohort Study (DECOS).

Zhu et al. adapt functional clustering methods to analyze discrete functional data generated from Electronic Medication Monitoring Devices (EMMD) in HIV prevention trials. Wisepill dispensers, an EMMD with built in cellular chips, provide objective and real-time measurement of adherence that is much more accurate than self-reported questionnaires. The authors discuss challenges for analyzing these regularly or irregularly spaced discrete functional data, and adapt three approaches for clustering including parametric, semiparametric and nonparametric approaches. Applying these methods to the HIV trial yields distinct temporal patterns of medication adherence.

Leroux et al. use the data from the NHANES study containing objectively measured physical activity collected using hip-worn accelerometers from multiple cohorts to develop a pipeline for data processing and dissemination. They accomplish it by providing an NHANES data package in R, to help disseminate high quality, processed activity data combined with mortality and demographic information. Using the developed tools, they show that accelerometry features have the potential to predict 5-year all-cause mortality better than known risk factors such as age, cigarette smoking, and various comorbidities.

Lee et al. study human gait based on the data from in-shoe sensors. They employ functional data analytic approach to detect gait phase and amplitude variability. They examine the correlation of phase shifts across sensors within a stance to evaluate the pattern of phase variability shared across sensors. They also explore associations between in-shoe sensor recordings and gold standard ground force reaction measurements to evaluate the in-shoe sensor recordings as their possible surrogate.

Krafty et al. study the daily sleep–wake pattern known as the rest-activity rhythm (RAR) and its association with health and well-being. They concentrate on the RAR measures quantifying variability around a mean circadian pattern and provide a new measure of RAR variability, namely the log-power spectrum of stochastic error around a circadian mean. This measure revealed that slow, rhythmic variations in activity from a circadian pattern are correlated with depression symptoms in a sample of adults suffering from depression.

Fadel et al. consider the problem of differentiating between walking and stair climbing using high-resolution raw accelerometry data. To classify walking and its subclasses, i.e., level walking, descending stairs and ascending stairs, they extract features from the raw accelerometry data based using a number of time- and frequency-based transformations and then build classification models using a tree-based methodology. The developed methods are applied to a sample of middle-aged subjects to evaluate their performance, including effects of sensor location and tuning parameters on classification accuracy.

Seewald et al. discuss micro-randomized trials (MRTs) which allow for data collection allowing construction of optimized just-in-time adaptive interventions (JITAIs). Such data are often collected passively using mobile phones. They next assess the causal effect of treatment on a near-term outcome and provide several recommendations for collecting and managing data from an MRT. Their recommendations are based on the HeartSteps study which was designed to test the effects of an intervention on increasing physical activity in sedentary adults.

Di et al. address statistical issues in the measurements of physical activity, quality of sleep, and strength of circadian rhythm arising from wearable devices. They propose to use joint and individual variation explained (JIVE), a dimension reduction technique that efficiently deals with multivariate data representing multiple domains to analyze the data obtained from participants of the Baltimore Longitudinal Study of Aging.

Xu et al. implement a functional principal component mixed model approach to study minute-level accelerometer multi-subject data. They apply their methodology to study temporal activity patterns in a study of overweight women. Summaries of individual patterns defined by personalized principal component scores are tested for associations with health outcomes including biomarkers such as insulin and C-reactive protein levels. Their model elucidates the most important patterns of variation in physical activities. They show that health outcomes are strongly associated with the total volume, as well as, temporal variation in activity.

Butera et al. are addressing missing data issues due to non-wear in epidemiological accelerometry studies. They study multiple imputation (MI) methods to handle missing accelerometry data and compare them on a simulated data with two more common methods of analyses, namely available cases (AC) and complete cases (CC) scenarios. They show that for the entire 24-h day, MI produced less bias and better coverage than AC and CC methods.

Kim et al. develop a modeling framework for the joint models of the longitudinal and survival data with no meaningful time-zero. They apply their methods to the longitudinal assessment of station during child labor and its relationship with the time to delivery. Their methodology needs to account for the competing risks. Their model is formulated through shared random effects between the survival and longitudinal processes, and parameter estimation is conducted through a Bayesian approach. The model is illustrated with longitudinal data on station and time to delivery.

Zhang et al. provides a review of statistical analysis for accelerometer-measured physical activity data, with a focus on studying temporal patterns of physical activity. They first review popular approaches of converting high-resolution movement data into to a variety of summary metrics, along with existing R packages for data processing. They then provide a comprehensive review of the longitudinal data and functional data approaches, including their multilevel and multivariate variations, for analyzing low-resolution summary data of physical activity recorded over short and long time periods.