Good afternoon everyone! My name is Xu Jing, my friends call me Kenny. Thanks for the department to organize the forum. It is my pleasure being here to share my research. As the first speaker, I would like say nice to meet you all.

My talk is about a novel micro randomized trial, used to study the effect of a developed mobile app that will send messages for encouraging physical activity. Before that, I would like to introduce my PhD and postdoc research.

I did a PhD of statistics from Macquarie University, Australia.

My topic was about fitting proportional hazard model under dependent censoring using maximum penalized likelihood. I had applied the proposed method to analyze a dementia dataset, and the data analysis results were accepted by Journal of Alzheimer’s disease. Based on the same dataset, I had completed a statistical methodology paper that published to SIM. We have an R package for the proposed method almost completed and planned to be submitted to CRAN.

The success of the methodology paper, I had received great support from my PhD supervisor associate professor Jun Ma and Professor Henry Brodaty, the director of dementia research centre at The University of New South Wales, Australia.

Comparing the existing likelihood method, our penalized likelihood method provides smooth baseline hazard estimates with consistent and asymptotic normality properties, under an efficient algorithm.

The future possibilities are, completing a software package paper, extending the proposed penalized likelihood method to fit other semi-parametric hazard models, applying to the precision medicine field, select the significant proteins for cancer survival outcome.

I did a postdoc of Biostatistics from Duke University in Singapore.

My first topic was about a novel version of sequential multiple assignment randomized trial for the treatment strategies of periodontitis. The second topic was about new version of Micro-randomized trial for the mobile device intervention. The second topic is ongoing and the main topic of this talk. For the first project, we have an statistical methodology paper accepted by Biometrical Journal and the corresponding R package is available at CRAN.

This is collaboration between statisticians and Dentist. For statistics part, it involves experts in both SMART design and Spatial modelling.

The novelties are, we proposed some novel treatment regimes for periodontitis, and a cluster-level of SMART design. Comparing to the existing SMART designs, the proposed design considers informative missingness, spatial associations with the cluster and non-normal outcomes. Again the R package of the proposed sample size calculator is available in CRAN. For the future research directions, the proposed sample size calculator can be extended to select the optimal treatment regimes, and work closely with the periodontist collaborator to run a study.

For the second project, is a type of data science collaborative research. It involves experts in statistics, social welfare and computer science. As the statisticians, we contribute experimental design, analysis plan and sample size calculation for the study proposed by the social welfare group, who developed a mobile application to encourage physical exercise. The algorithm used to support the mobile application are developed by computer science group. This talk focuses on the statistical part.

I will start a background introduction of mobile health, then the real study. The main parts are the proposed design, analysis plan and sample size calculation. The simulation studies will be given to validate the performance of the proposed sample size calculator. The presentation will be ended by the discussion.

First, I will give a brief background introduction for the mobile health.

What is mobile health? It is the practice of medicine and public health supported by mobile or portable devices. It is affordable and user friendly.

The motivation of mHealth development is mobile phone becomes more important to our daily life.

Especially for the developing countries, where got the constraints for the healthcare support.

The strength of mHealth interventions can be best realized when they can **adaptively** respond to individuals’ actions and states and deliver intervention options that are most needed **whenever and wherever**, we called that just-in-time adaptive interventions.

Here, we list some example of mHealth utilizations, which are improving HIV medication, increase physical activity, supplement counselling or pharmacotherapy in treatment for substance use, and alcohol or smoking cessation.

Next, I will briefly describe the study proposed by the social welfare group.

Both the study and mobile app are named DIAMANTE. It stands for diabetes and mental health adaptive notification tracking and evaluation.

The overall goal of the study is to develop, implement and evaluate the DIAMANTE app. This app can be used as a mobile intervention on physical activity. The intervention will improved through adaptive learning. The cohort of the study focuses on the diabetes and depression patients with low income.

The intervention is sending messages through the app. Physical activity is walking. Adaptive learning is to continuously improve the messages to participant to encourage them walk more. Messages are multi-component with varying level. There are three components, which are time window, feedback message and motivational messages.

Time window considers when to send the message during a day within study period, such that morning, mid-day, early afternoon or early evening, which is four levels.

Feedback message is a five-level intervention. The base level is control, no message. We are interested to test if one of the four interventional level of motivational message is effective on walking. Those levels are, feedback to participants, whether reached the goal, calculation the number of steps walked yesterday, telling that if walked more or less than yesterday, telling the steps walked yesterday and a adding a encouraging message, for example, great jobs.

Similarly, we proposed a four levels motivational message. The interventional levels focus on benefit, self-efficacy, and opportunity of walking.

Now, I will talk about the proposed study design.

Since DIAMANTE is multi-component, which can be studied by the factorial design. But factorial design does not consider time-vary and sequential nature of intervention, which can be captured by a micro-randomized trial design. It is sequential full factorial design that captures the just in time intervention purpose of mobile intervention. Each participant is randomized for the possible action at each decision time point. Outcome is proximal, the number of steps walked in the next 30 mins after a message is sent. Since actions are time varied, the data structure is longitudinal, generalized estimating equation type of analysis will be employed.

The proposed design is called multi-level micro randomized trial design. It considers components with more than 2 levels. In addition, sample size calculation is not only can be based on power but also precision. For each component, adding new levels is possible for sample size calculation. Sending better messages to encourage participants to walk more through online learning strategy.

After talking about the study design, I will move to the proposed statistical analysis plan and sample size calculation.

For the data structure, each participant will be observed at T decision time points. At each time point, we will observe the pre-treatment information, treatment category, proximal outcome, i.e. steps walk within next 30 mins.

The proximal effect of each interventional messaging level, at a particular time t, can be defined by the difference of the expected number steps taken within next 30 minutes between the interventional messaging level m and the control level.

The null hypothesis is no proximal effect for each treatment category at each decision time point. The alternative hypothesis is that the proximal effect can be model by a polynomial function. Betas are the parameter interested.

The parameters are estimated by least square method based on the availability of each participant at each decision time point. We have a theorem of consistent and asymptotic normal properties for the least square estimator. Therefore the square estimator times the inversed of the asymptotic covariance estimator follows chi square distribution.

For large sample, the test statistics follows a chi-square distribution. It is central distribution if null hypothesis is correct, otherwise it is non-central distribution. For small sample, Sigma is replaced by Sigma hat. The test statistics follows a hoteling t-square distribution, which is in proportion of F-distribution.

Base on the distribution of the test statistics, the required sample size can be calculated by either power or precision based method. For the power based method requires inputs of number of decision time points, effect size, desired power and level of significance. For the precision based method, power is replaced by coverage probability and effect size is replaced by the margin of error.

In the following, I will demonstrate the performance and application of the sample size calculator through some simulation studies.

We specify the power to be 80%, 5% level of significance, number of decision time points of one per day, small effect size, which increase until 28 days and maintain the max value for the rest. We assume 100% availability.

Given all the categories of each component are proposed at the beginning, for power based sample size calculation, we observed that the estimated sample sizes give both the powers of formula based on Monte Carlo based are close.

For the precision based sample size calculation, the corresponding coverage probability of both formula based and Monte Carlo based are close.

The similar results are observed for the component categories proposed both at the beginning and middle way.

For the discussion part, I will summarize some key points and propose some potential future research directions.

For the summary, mobile or digital treatment can stabilize human health from global perspective. Multi-Level Micro-Randomized Trial is a cutting edge trial design in mobile or digital health. The sample size calculator will be available in public soon.

This is my schedule for the ongoing work. In one year time, I will run more simulation studies, and complete the manuscript and aim to submit it to a high rank statistical journal, as well as the corresponding R package and R shiny app. In three years time, I will provide statistical analysis for DIAMANTE study, and look for methodology research, complete a paper for the R package and submit to a statistical software journal. In five years time, I will look for other mHealth domain collaborative opportunities.

This research work is financially supported by Singapore parts are Duke university, institute of mathematical science, national university of Singapore, ministry of education and US part, national institute of health.

For further discussion about the slides, please feel free to contact me through my duke-nus email.

Again, I would like to thank for your attention and all the support from the department of statistics and data science. Now I am looking forward for your feedback and ready for the rest of talks.