Breakout Session 8: Track B

AI/ML Ready mHealth and Wearables Data for Dyadic HCT

Dr. Vibhuti Gupta
Assistant Professor, School of Applied Computational Sciences, Meharry Medical College

AI/ML Ready mHealth and Wearables Data for Dyadic HCT

Rajnish Kumar PhD

Postdoctoral Fellow

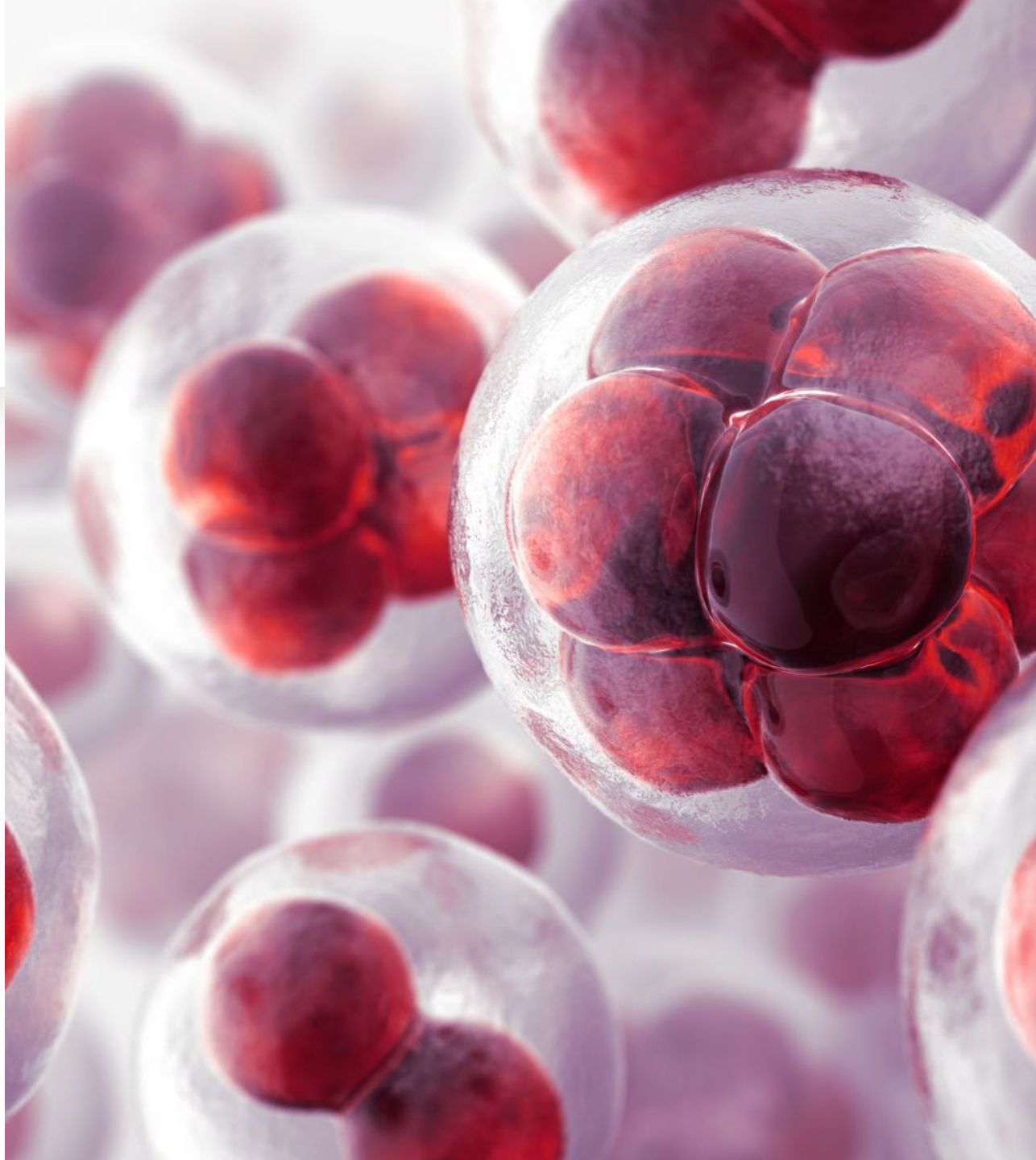
Choi Lab | University of Michigan

Blood and Marrow Transplantation Program

Vibhuti Gupta PhD

Assistant Professor

Meharry Medical College



*This work is supported by an NHLBI Supplement Award – **K24HL156896***

Talk Outline

- mHealth Study Overview
- Research aims
- Preprocessing Pipeline
- Results
- Challenges

Project Objective

- To utilize pre-processing procedures and develop novel methods to enrich the quality of data, captured from smartphone sensors and wearable devices to make it ready for AI/ML applications.

- Share the processed data and developed methods through repositories, and knowledge bases to enhance the ethical re-use of data for AI/ML applications.

Study Population: Hematopoietic Cell Transplantation

Hematopoietic cell transplantation is an important therapeutic option for a number of blood diseases.

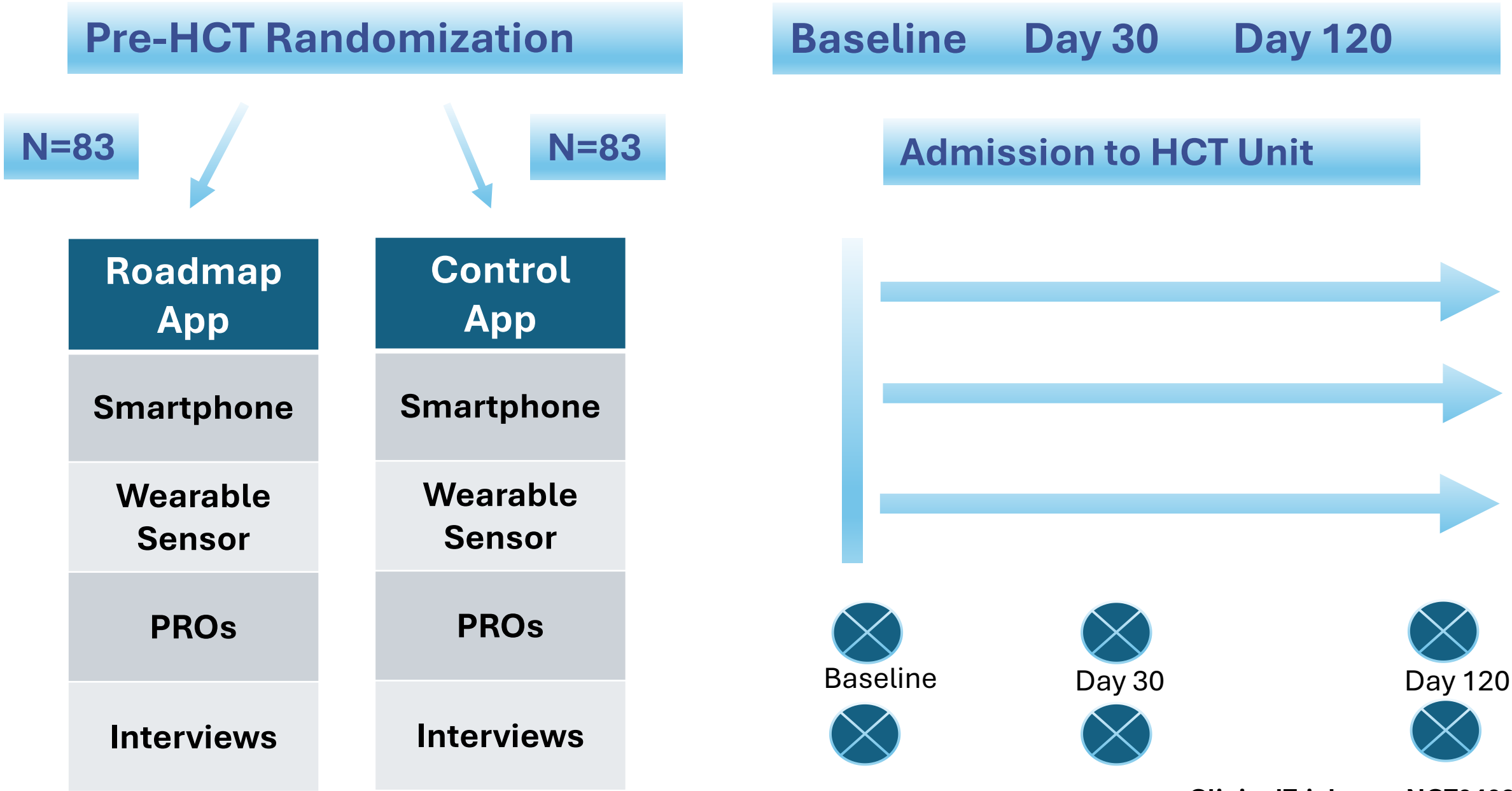
The treatment is intense and requires a dedicated 24/7 care partner for at least the first ~100-days of treatment.

Care partners are often considered “hidden patients” given the immense burden that they carry (i.e., managing medications, appointments).

It is not uncommon for care partners themselves to experience adverse physical and mental health.

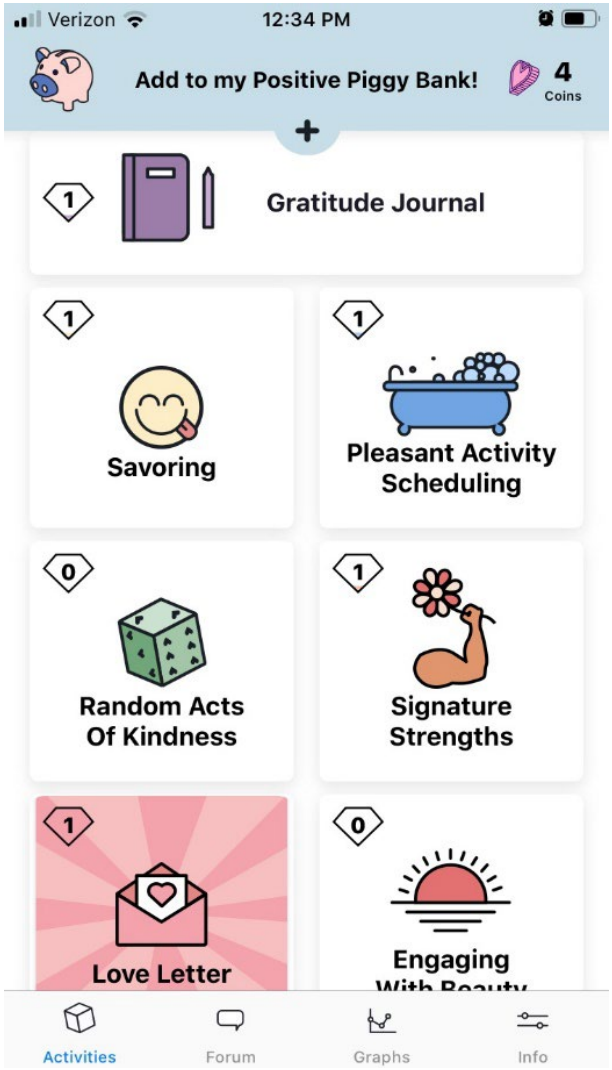
Roadmap app

- Choi Team developed a multi-component app, called **BMT Roadmap,** targeting care partners and their patients
- Multi-component app compatible with iOS and Android:
  - Positive Activity Features
  - Chat Forum
  - Graphs (mood, sleep, steps–past week data)
  - Integrates with Fitbit Charge devices

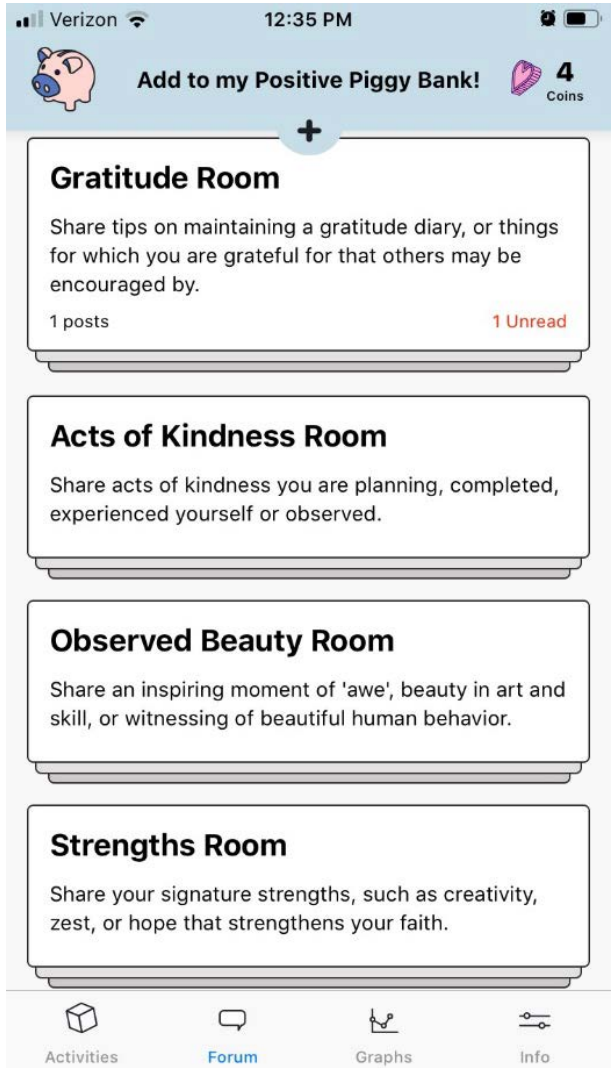


ClinicalTrials.gov NCT04094844

BMT Roadmap

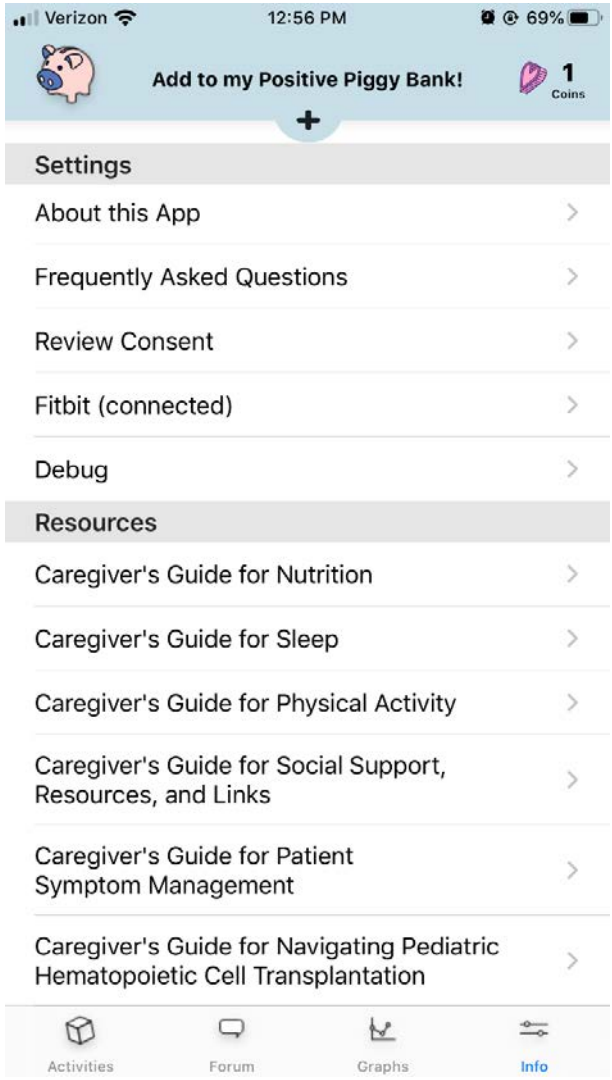


Positive Activities



Chat Forum

Fitness tracker + daily mood



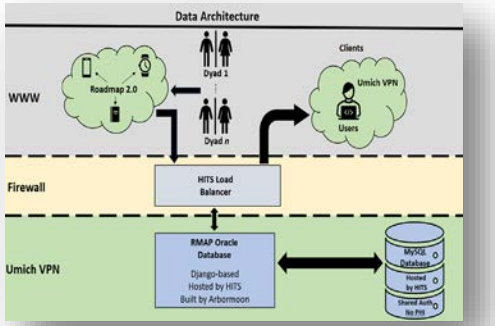
Resources

Specific Aims

- Aim 1: To develop high-quality mHealth datasets for AI/ML applications in HCT research. Activities to support this aim include:
    - Pre-processing procedures and developing automated and scalable methods for efficient data cleaning
    - Integrating data from disparate sources to make it ready for applying AI/ML techniques for required project outcomes
    - Developing novel techniques to reduce biases in the data for advancing the ethical development of AI-ready data

Overall Preprocessing Pipeline

Data Extraction → **Data Standardization** → **Data Cleaning** → **Feature Extraction** → **Data Integration** → **AI/ML ready data** → **Ethical development of AI/ML-ready data**



- Duplicates removal
- Checking data validity
- Compliance algorithms to filter data

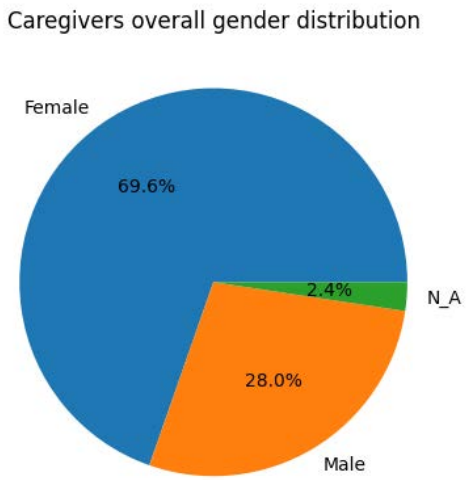
- Data Smoothing
- Detection and removal of outliers and inconsistent data
- Data Normalization
- Missing data imputation

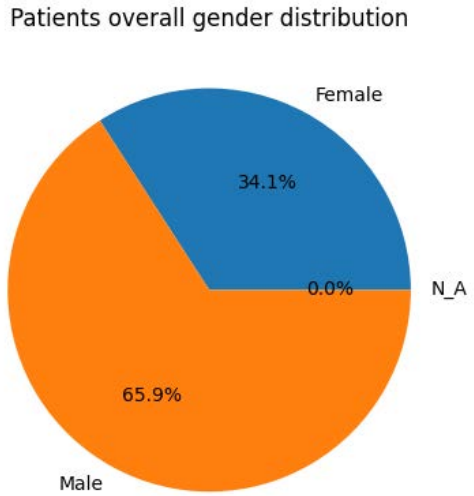
- Statistical Summaries
- Visualization
- Window-based feature extraction

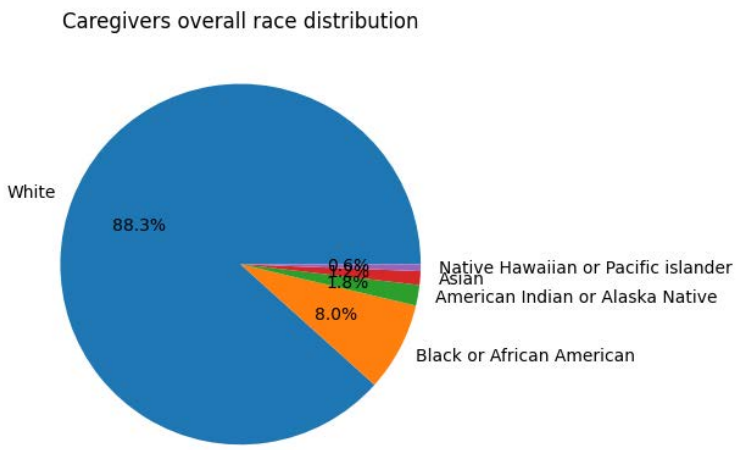
- Resampling
- Data augmentation

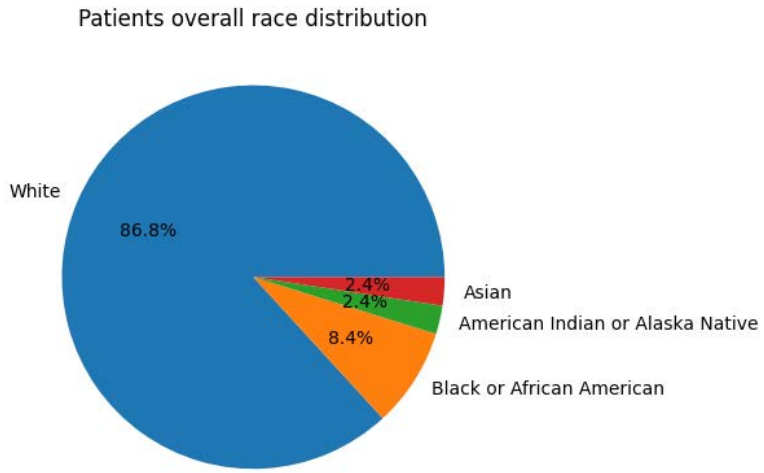
- Early Fusion
- Late fusion

Data Description and Demographics









- Data collected from Michigan Medicine from an existing mHealth randomized clinical trial (NCT040984844)
  - **September 2020 – July 2023**
  - **166 Caregivers and 166 patients' data**
- Fitbit data
  - Heart rate -
  - Sleep
  - Steps
- Data from Roadmap App
  - Survey
  - Patient-reported outcomes (baseline, daily, monthly)
  - Mood

Fitbit Data

- We observe **506,517,679** observations across all 323 participants
  - **Steps:** 417,003,290 observations across all 323 participants
  - **Sleep**: 3,075,687 observations across all 323 participants
  - **Mood**: 381,164 observations across all 323 participants
  - **Heartrate**: 86,057,538 observations across all 323 participants

Highlights on data preprocessing

- **Data Duplicates**
  - A total of 45.67 % of the data is duplicate data without any statistical replacement in the data

Steps

- A total of 48.3 % are duplicates

Sleep

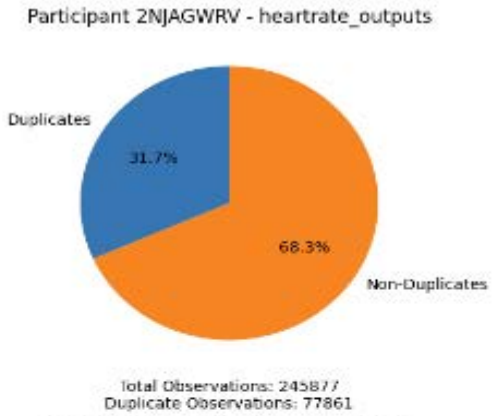
- A total of 0% are duplicates

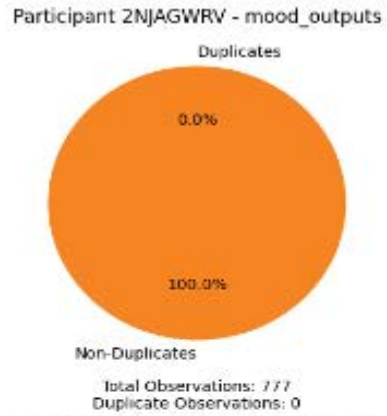
Mood

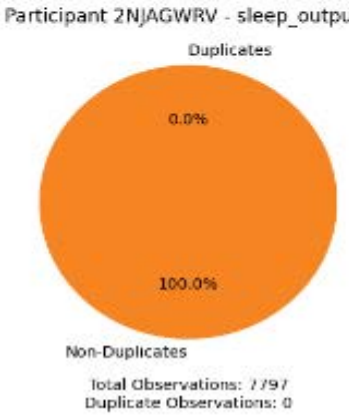
- A total of 0.01% are duplicates

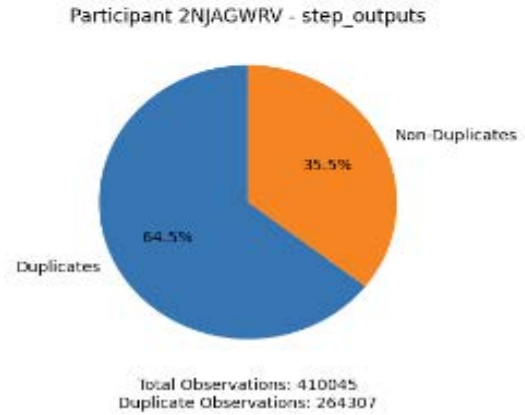
Heartrate

- A total of 34.55% are duplicates









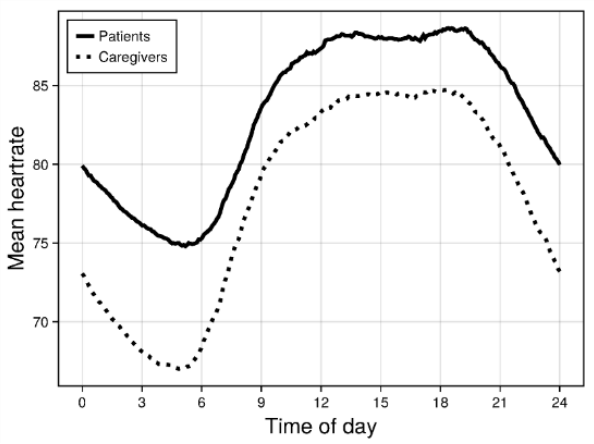
Highlights on data preprocessing

- **Data Standardization**
  - Developed compliance algorithms to filter the data based on data validity and compliance
  - Applied to average daily step count and average daily heart rate data

  Metrics for compliance
    - Wear time: Ratio of the number of minutes registered heart rate and total minutes for a day
    - Validity: A valid day corresponded to a day that met a certain criterion and was kept for analysis. Three definitions:
      - (None) all days were considered valid
      - (StepCount1000) a day was valid if the step count registered for that day was greater than 1,000
      - (WearTime80) a day was valid if the wear time that day (24 hrs) was greater than 80%

Baroudi L, Zernicke RF, Tewari M, Carlozzi NE, Choi SW, Cain SM, Using Wear Time for the Analysis of Consumer-Grade Wearables' Data: A Case Study Using Fitbit Data, JMIR Preprints. 31/01/2023:46149

Some preliminary dyadic Heart rate observation



Mean HR trajectories for patients and caregivers

- Factor analysis was done to identify the strongest dyadic components while preserving individual explained variation.
- Objective function included that of canonical correlation analysis along with principal component analysis of patients and caregiver groups separately.

bioengineering

New Publication

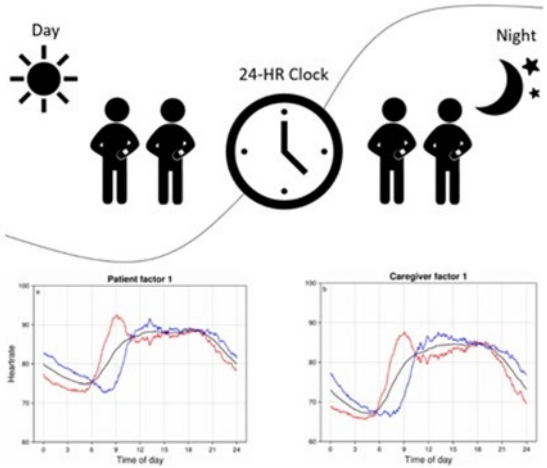
IMPACT FACTOR 4.6

Indexed in: PubMed

Dyadic and Individual Variation in 24-Hour Heart Rates of Cancer Patients and Their Caregivers

Authors: Rajnish Kumar, Junhan Fu, Bengie L. Ortiz, Xiao Cao, Kerby Shedden and Sung Won Choi

Bioengineering **2024**, *11(1), 95*; https://doi.org/10.3390/bioengineering1101009



Key points:

- ❖ Heart rate data from cancer patients and their caregivers uncovered unique patterns with link to demographic, environmental, psychological, and clinical variables.
- ❖ Five factors of dyadic and individual heart rate patterns were identified using dimensionality reduction approaches.
- ❖ Interestingly, higher mood in cancer patients corresponded with earlier heart rate decrease in the morning and increase during the afternoon.

Bioengineering, ISSN 2306-5354, IF 4.6 Rapid Publication

Submitted to Bioengineering

Specific Aims

- Aim 2: To share the processed data and developed methods for AI/ML and mHealth research community. Activities to support this aim include:
    - Documentation for the processed data
        - Developing data dictionary for the processed data
    - Processed data is shared through secure data repositories and knowledge bases
        - Data and methods will be released through Michigan Deep blue repository and Github

Challenges

- High dimensional data
- Frequent and irregular data sparsity
- Compliance issues
- Multiple device IDs for several participants

Conclusion and Future Work

- The wearables data is messy, high-dimensional, and requires a lot of effort to preprocess and enrich the data quality.
- We are currently working on developing the methods to preprocess data
- A systematic review paper is under review to list the preprocessing techniques used to improve the data quality in mHealth studies in cancer.
- An empirical data analytics paper showing our findings is in progress
- We will apply AI/ML techniques once data is ready to understand the mental health behavior using physiological parameters from wearables.

Acknowledgements

PI* & Co-Investigators	Postdoctoral Fellows	Research Coordinators
Sung Won Choi MD MS*	**Rajnish Kumar PhD**	Uma Subrayan MS
Vibhuti Gupta PhD	Bengie Ortiz Del Valle PhD	Michaela Carter BS
Debra Barton PhD RN	Jonathan Tyler PhD	Caroline Clingan BS
Tom Braun PhD	Chris Flora PhD	Christine Cislo BS
Noelle Carlozzi PhD	Amanda Johnson MD	**Funding Sources**
David Hanauer MD MS	**Graduate Students**	NHLBI (R01HL146354)
Afton Hassett PsyD	Caleb Mayer PhD	AHRQ (R21HS023613)
Institutional Support	Jiyoun Shin PhD	NHLBI (K24HL156896)
Rogel Cancer Center	**Biostatistician**	Edith S Briskin/ Shirley K Schlafer Research Professorship & Taubman Medical Research Institute
BMT Program	Xiao Cao MS	
Thank you to all of our patients and caregivers!	**Data Manager**	
	Michelle Rozwadowski BS	