2016 All-Hands Annual Meeting Report
Held October 2, 2016 & October 3, 2016
University of Memphis

Grant Number: 5U54EB020404
Program Director/Principal Investigator Information: SANTOSH KUMAR, PHD

MD2K YEAR THREE PLANS

MD2K Hands-On Software Training Day
FedEx Institute of Technology - Memphis, TN
October 3, 2016
- Course Description -

The MD2K Hands-On Software Training course provides a hands-on introduction to our smartphone platform (mCerebrum) and our cloud platform (Cerebral Cortex). The day will begin with every participant receiving a set of hardware to wear and going through a tutorial on putting everything on and configuring mCerebrum to collect, process, and upload data. The remainder of the day will continue to utilize the ongoing data collection to demonstrate and explain components of the platforms.

Subsequent sessions will provide more in-depth discussion of various components of mCerebrum and Cerebral Cortex suitable for both the health-science and data-science communities. It will highlight key components including: DataKit, a public data bus capable of transporting 3 million samples per hour, how we achieve real-time marker computation, various significant data streams with details to specific algorithms presented by their respective authors, and the various human interface mechanisms we support and enable with mCerebrum.

The afternoon will explore how Cerebral Cortex's interface can be utilized to monitor ongoing studies and participants, how to export data from the cloud system for analysis in a variety of tools, the specifics of how data is inserted into the platform, some of the back-end data processing mechanisms including a data diagnostics tool designed to identify failures from analyzing various data streams, and finally how we train our models for the cStress and puffMarker algorithms.

The day will conclude with an overview of the future directions we wish to take the platform including new capabilities and features and a discussion of how one can contribute to the system.

Participants who take this training will receive a certificate of completion

MD2K Hands-On Software Training Day
FedEx Institute of Technology - Memphis, TN
October 3, 2016
– Agenda –

[9:00 am -10:00 am]
Session 1: ***Introduction to mCerebrum and Cerebral Cortex***

a. Welcome and overview - *Tim Hnat*
   i. Overview of the Day
      1. Learning Objectives: What do they get out of the session?
      2. Training Overview
   ii. Software Video
b. Hands on session (Using mCerebrum and Cerebral Cortex) - *Shahin Samiei, Brian Ahern*
      1. Handout devices
      2. Configure and setup mCerebrum - Everyone
      3. Visualize data (Plotter)
      4. Verify data collection and upload - Everyone supported by Memphis team

[10:00 am - 10:30 am]
Session 2: ***What is in it? (mCerebrum - the smartphone platform)***

a. Architecture overview and features
   i. Layered and modular design
   ii. DataKit - A public data bus transporting 3 million samples per hour
      1. Pub/Sub to support data and computation reuse
      2. Built-in support for privacy control
      3. Data offloading and archiving
   iii. Real-time marker computation - Computations must not fall behind
      1. Pub/Sub to support data and computation reuse
      2. Built-in support for privacy control
   iv. Data offloading and archiving
   v. Significant data streams
      1. High-frequency raw data: ECG, respiration, accelerometer, gyrometer, battery, GPS
      2. Features: RR-interval, heart rate, HRV, respiration rate, inspiration and expiration durations, and 30+ additional features
      3. Momentary biomarkers: stress probability, smoking puffs, driving, activity
      4. Episodes: Stress, Smoking

[10:30 am - 11:00 am]
Break

[11:00 am - 12:30 pm]
Session 3: ***How does it work? (mCerebrum)***

a. Data sources and Communication interfaces:
    i. Phone Sensors, Microsoft Band, Omron BP and Weight, AutoSense, EasySense, MotionSense, Oral-B
b. Signal processing and Biomarker
    1. RIP
    2. ECG
    3. Wrist
    4. cStress
    5. puffMarker
    6. driving/riding
    7. data quality
c. Participant User Interface
d. User Input and Prompting
e. Signal processing and Biomarker

[12:30 pm -2:00 pm]
Lunch and Networking

[2:00 pm - 3:00 pm]
Session 4: Cerebral Cortex Architecture and Capabilities

a. Data collection monitoring
    i. Study Dashboard (demo w/ real data)
    ii. Data Visualization (demo w/real data)
b. Data export
    i. Demonstrations of CC export capabilities (demo w/real data)
    ii. Examples for R, Matlab (demo w/real data)
c. Data processing
    i. Data ingestion
    ii. Data diagnostics
d. Model training

[3:00pm - 3:30 pm]
Session 5: Future directions

a. mCerebrum
    i. iOS porting
    ii. New biomarkers: Eating, Speaking, Brushing, TV Ad exposure
    iii. New sensors: Eyeglasses, Android wear/smartwatches, commercial chest bands
    iv. Security and privacy: Access Controller, Encryption-based Privacy wrapper, Smart obfuscation
b. Cerebral Cortex
    i. Discovery Dashboard
    ii. Unified data store (AsterixDB)
c. AWS production environment
    i. Overview background of cloud computing?
d. Open mHealth integration

[3:30 pm - 4:00 pm]
Session 6: How do I contribute?

a. Code b. Testing

Conclusion

Third Annual MD2K Meeting
FedEx Institute of Technology - Memphis, TN
October 4-5, 2016
- Agenda -

Tuesday, October 4, 2016

7:00 – 8:00 **Breakfast**, University of Memphis Holiday Inn – Medallion Restaurant
8:30 – 8:45 Welcome from Center Director, NIH, Team Introductions - *Santosh Kumar*
8:45 – 9:00 Introduction to Data Science Research Core – *Jim Rehg*
9:00 – 9:20 Sensor Platforms – *Emre Ertin*
9:20 – 9:40 Sensors to Markers – *Deepak Ganesan*
9:40 – 10:00 Markers to Predictors (Predictive Analytics) – *Jim Rehg*

10:00 – 10:20 **Break**

10:20 – 10:40 MD2K Computational Platforms & Architecture – *Mani Srivastava*
10:40 – 11:00 MD2K Software Platforms – *Tim Hnat*
11:00 – 11:20 MD2K Standardization Activities – *Ida Sim*
11:20 – 12:00 Sensor and Software Development Process - A Historical Perspective – *Emre & Tim*

12:00 – 12:30 **Lunch & Networking** (FedEx Institute of Technology)
12:30 – 1:30 MD2K Poster and Demo Session (for the MD2K Team)
1:15 – 1:30** ***MD2K Team Picture*

1:30 – 2:00 MD2K Smoking Study - Current Status & Future Plans – *Bonnie Spring*
2:00 – 2:20 MD2K CHF Study - Current Status – *William Abraham*
2:20 – 3:00 Brainstorming - Publications plans & Second Iteration of the Study – *William Abraham*
3:00 – 3:20 MD2K Participant Engagement Efforts & Plans – *Jim Rehg*

3:20 – 3:30 **Break**

3:40 – 4:00 Training Core - *Vivek Shetty*
4:00 – 5:00 Concurrent Breakout Sessions:

a) MD2K Future Funding/Sustainability Planning – *Santosh Kumar*
in the Methodist Theater – FIT 103

b) Students to meet with NIH – *Richard Conroy*
in the Digital Living Room – FIT 315

5:00 – 5:30 **Break**
5:30 – 6:30 Reception, Networking, and MD2K Poster and Demo Exhibition (*public event*)
6:30 – 9:00 **Dinner**, FedEx Institute of Technology

Third Annual MD2K Meeting
FedEx Institute of Technology - Memphis, TN
October 4-5, 2016
- Agenda -

Wednesday, October 5, 2016

7:30 – 8:30 **Breakfast**, University of Memphis Holiday Inn – Medallion Restaurant

9:00 – 9:30 Consortium Core - *Ida Sim*

9:30 – 10:20 Data Description, Sharing, Access, and IRB - *Shahin Samiei*

10:20 – 10:40 Administration Core – *Joe Biggers*

10:40 – 11:10 MD2K Process and Reflection - *Santosh Kumar*

11:10 – 11:50 MD2K Impact - *Santosh Kumar*
How to enhance MD2K impact with dissemination and adoption of our science, sensors, software, and training?

11:50 – 12:00 Conclusion - *Santosh Kumar*

12:00 Box Lunch

MD2K Annual Meeting concludes

1. Introduction

This is a summary of the 2016 Annual Meeting. More than 70 members of the team came to the University of Memphis in October for a 3-day event that began with a day of hands-on training with MD2K's software and also featured updates from MD2K research teams, student poster presentations and demos and lots of discussion about future plans for MD2K as it enters its third year of work.

This report summarizes accomplishments and goals for each research area as well as updates and other news.

For the purposes of this report, Year 1 refers to 9/2014-9/2015, Year 2 refers to 9/2015-9/2016, and Year 3 refers to 9/2016-9/2017.

1.1 Overall Accomplishments

- MD2K researchers have had 200+ articles accepted for publication in 2 years.
- James Rehg, Santosh Kumar and Susan Murphy are editors on a book that has been submitted to Springer for publication.
- 20 apps have been developed for mCerebrum (3 are intervention apps) + Cloud.
- Emre Ertin's sensors (Autosense, MotionSense, EasySense) are being widely used.
- Studies: Bill Abraham's pilot study for Congestive Heart Failure has been approved by IRB; and Bonnie Spring has completed a groundbreaking study design (eating & smoking with just-in-time intervention).
- Vivek Shetty's mHealth Training Institutes: 2016 completed, work on 2017 has begun. So far training has been given to 70 scholars from 50 institutions plus almost 4,000 views of the mHTI videos online.
- A total of 60 lectures, 45 articles are available on the mHealthHUB, which has had more than 62,000 page views.

2. MD2K Data Science Research (James M. Rehg)

Brief overview of activities in DSR Core

The Data Science Research (DSR) Core is tasked with finding ways that mobile sensor data can be collected, processed and stored in a way that enables its use by the broader mHealth community. In an overview of the DSR core, Dr. James M. Rehg, deputy director of MD2K and a professor and researcher at Georgia Institute of Technology, outlined the changes and accomplishments of each Thrust:

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

There were a few changes in the organization of Thrust 2 this year with the creation of a Predictive Analytics component and a Participant Engagement component focused on how to meaningfully integrate participant engagement into the studies.

Thrust 3 was also reorganized into components focused on Mechanisms and Architectures and Standards.

MD2K has really "hit its stride" this year and engaged some of the problems that the center has been tasked to investigate.

2.1 Thrust 1:

Accomplishments:

- Markers for stress, smoking and eating have been improved and incorporated in the design of the user studies.
- A novel marker for craving has been developed.
- EasySense has made great progress.
- Vision-based sensing (e.g., cameras) is progressing as well.

Plans:

- Additional marker development (novel markers from respiration).
- Use wearable cameras more aggressively.
- Increase use of EasySense.
- Integrate geoexposure.

Thrust 1 has made a strong investment in methodology. Despite the importance of sensors, the group has also been mindful of improving the utility/methodology of the sensors.

2.2 Thrust 2:

Year 2:

Accomplishments:

- A pipeline has been established Just-in-Time Intervention from data streams to markers.
- Progress has been made on how to interact/intervene on predictors of negative events.
- The Discovery Dashboard system was released to the Thrust 2 group several weeks ago.
- Stress marker is now being used in studies.

Plans:

- Continue to work toward improving stress predictors.
- Continue to work on improving Just-in-Time Interventions.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

2.3 Thrust 3:

Accomplishments:

- Adoption of software at six different sites.
- Use of back-end platform (Cerebral Cortex).

Plans:

- Work on privacy control and standards.

2.4 Thrust 4:

Accomplishments:

- Hardware has been incorporated into studies.
- Software is being deployed to the field setting.
- Are integrating the micro-randomization component.
- Launch of the CHF pilot study is a tremendous accomplishment.

Plans:

- Enrollment in field study planned for October 10.

3. Sensor Platforms (Emre Ertin)

Overall goal is to provide a plethora of sensors and push data.

3.1 Sensors currently being deployed:

3.1.1 Autosense — Physiological-based sensors that stream ECG and RIP data to phone. Autosense streams ECG and RIP data to phone.

- The respiration circuitry has been revised to improve detection of smoking through signal processing.
- A new "clip" to improve ergonomics has been added.
- An upcoming Bluetooth improvement will allow the user to select which sensors to energize.

3.1.2 MS Band — Contains a display screen for prompts, PPG sensors for heart rate. Unfortunately, MSBand is being phased out, so the team is looking for replacement.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

3.1.3 Easysense — EasySense measures lung fluid in CHF studies: A microdata platform sends pulses to body and records backscatter to measure changes in tissues, monitor the motion of heart and lungs to capture respiration signal. It's a kind of "poor man's ultrasound," with multiple antennas and receivers.

- EasySense gathered a lot of info during pilot and is now in validation stage
- It is a fully functional sensor system with GUI, the patients did not like the front-and-back sensors so we moved to backscatter approach, which is an easier path to commercialization.
- Early work resulted in too much variance of data, so a new calibration channel was added to improve signals/data and moved from rectangle to circular.
- Successfully decreased variance, can be worn over the clothes or on bare skin. Under consideration is giving participants a t-shirt with a mark on it to improve consistency in sensor placement
- It provides 1-minute measurement and is a standalone unit (no longer requires laptop), so all data is stored on the device's SD card.
- It has integrated ECG sensors to measure arterial stiffness, 10cm depth into body.

3.1.4 Eyeglasses — Developed by Deepak Ganesan's team for detection of pupil dilation and saccadic movement.

- It requires being inventive in signal processing.
- The team has improved ergonomics and look of design.
- Integrated imagers into glasses frame.

3.1.5 Motionsense wrist band — This custom band is geared towards gesture recognition.

- 9-axis data stream.
- Improved appearance.
- Bluetooth low energy, wireless charging.
- Waterproof.
- Improved battery life.
- State-of-the-art monitor to detect driving, etc.
- Sensor has been integrated into mCerebrum, 5000 capable IMU.
- The chip has built-in motion processing for gait count, etc.
- The firmware is on Github and is open source.

3.1.6 Smartphone sensors — Gyroscope, accelerometer.

3.1.7 Commercial sensors — such as Omron scale and blood pressure cuff and JINS eyeglasses.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

4. Thrust 1: Sensors to Markers (Deepak Ganesan)

4.1 Marker development and validation

- Markers of interest that have been identified for biomedical applications for CHF and smoking cessation:
    - lung fluid
    - fatigue
    - eating habits
    - smoking habits
    - craving
    - stress
    - exposure
- Methods being used to sense these markers:
    - Easysense (micro-radar
    - iShadow/Pivothead (imager)
    - Earphones (infrared, gyroscope)
    - AutoSense (ECG, respiration)
    - Smartphone (GPS, self-report)

The goal is to achieve high precision and accuracy of marker detection in an unobtrusive sensor modality. There are questions:

- Sparse-compressive sampling for power optimization
- Feature engineering to extract discriminative features
- Models that are robust to significant situational variability
- Learning time-series detectors despite errors in self-report data
- Learning Robust models with small amounts of labeled data.

4.1.1 Measurements of lung fluid (Gao, Ertin, Ohio State):

- Goal: Detect minute changes in the body's dielectric properties due to lung water accumulation (via Easysense)
- Developed a statistical feature for detection of lung water which has had promising preliminary results in a pilot study.

4.1.2 Fatigue sensing with the iShadow platform (Rostaminia, Ganesan, Marlin, et al, UMass Amherst):

- Estimates of blinks and eye-closure patterns to associate with levels of fatigue or drowsiness.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

  - Bling detection 95% precision; 85% recall
  - Blink duration 97.6% accuracy
  - Percentage of eye closure: 97.5% accuracy
  - Power consumption: ~45mW @ 100Hz
- Importance: Drowsiness is a safety-critical phenomenon.
- Sampling at 100Hz to extract markers of interest.
- Results on 16 subjects has been promising with high precision and recall.

JINS Electro-ocular measurements:

- One-dimensional signal. Easier to see the "hard/full blinks" -- harder to see the finer-grain movements such as saccades.
- Electro-ocular measurements considered a gold standard with sufficient sensors (electrodes).

4.1.3 Eating habits (Thomaz, Abowd et al, Georgia Tech):

- Tracked with smartwatches and earphones.
- Many wrist-worn devices have accelerometers.
- No issues running accelerometers for long periods, but with what utility for behavior detection?
- A pilot study of 21 participants, 75-80% accuracy.
- Inferential power of dominant vs. non-dominant hand.
- Year 3: Classify over-eating vs. emotional eating; oral health behaviors (brushing, flossing).

Eating detection using earphones:

- Outer ear interface -- embedding sensors into an earpiece to detect chewing: Three proximity sensors in different angles that detect deformations in the ear canal.
- Using gyro to help remove confounders (e.g., walking). Recall very high in lab (93%) and field (95.4%), more work needed on precision in the field (~60%).

4.1.4 GPS (Kerr, Raab, Patrick et al, UCSD):

- Positive predictive value of GPS is riddled with noise. (Did not eat, purchased food to go, were in the car the office or nearby retail)
- More work needed to improve reliability of GPS/GIS information.
- Proposing a probability layer that smooths risk estimates and can include contextual factors.

4.1.5 Craving (Chatterjee, Hovsepian, Kumar et al, Memphis):

Development of computational model to estimate craving each minute during smoking abstinence. Can we turn craving into a continuous measure from EMA & mobile sensor data? Predicting craving over the time of day: Using stress and the time of day is much more accurate than using either alone.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

4.1.6 Exposure (Rehg et al, Georgia Tech):

Creating an objective measurement of TV-watching behavior using iShadow/Pivothead to analyze content to identify TV ads. A large trigger for smoking relapse is exposure to contextual factors (e.g., alcohol advertisements).

The work is looking at ad exposure from first person vision. Is the viewer actually looking at the TV, at a TV alcohol ad, that may trigger craving? High accuracy of detecting a bottle in the ad.

4.1.7 Hierarchical Segmentation and Labeling for Time Series (Adams, Thomaz, Kumar, Marlin, et al)

How to segment data streams to run a classifier? ECG is very structured, others like accelerometry aren't so much. The work is determining how to segment events into high-level activity sessions (e.g., smoking episodes or eating sessions). Promising benefits from work with pre-MD2K datasets.

4.1.8 Learning with Temporally Noisy Labels (Adams, Marlin, UMass Amherst)

The challenge in using self-reporting is how to address the disparity between physiological signals (sub-second timescale) and rate of human response (minute timescale) -- these windows often do not overlap, introducing mathematical uncertainty.

The team is working on full model of this temporal uncertainty to account for introduced noise.

4.1.9 ECG Feature Detection Using Random Compression (Gao, Teng, Ertin, Ohio State)

Sparse/compressive sampling: Knowledge that random sampling can yield high quality data (e.g., in fMRI). Gao et al. discovered that compressing sampling by 30x still yielded high fidelity data signal.

Validation: What validation will or can work in the field setting? In the field, self-report is one of the few validation steps that we can get (with its own accuracy issues). Cameras are another method. Some markers are objective (eating can be viewed with a camera). Others (craving, stress) are harder to objectively validate, detect or measure. Jim: Sometimes this feels like a collection of case studies rather than an overarching framework. Mustafa: There is a process for setting criteria -> validation. Ida: Taking established methodologies and applying them to areas that aren't as established. Bonnie: Communities (e.g. tech and clinical) have had to dialogue to compel movement in the advancement of using different sensing methodologies. Ida: There are different thresholds for validation: there is validation for different purposes. Bonnie: Shaping the consumer perception of validation -- what is sufficient? Karen: Hierarchical self-reports that address interval-level data (rather than punctate) can also be informative. Use self-reports of self-reports to model the accuracy of data gleaned (e.g., from EMA). Gregory: It all comes back to ground truth. Billie: There is no unified definition for "fuzzy" concepts like stress -- we need decisions regarding definitions.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Markers	Sensors	Status/Investigators
Markers relevant to smoking and eating		
Detecting of smoking from respiration and IMU data	AutoSense	Kumar, Ertin,Marlin [Ubicomp 2015]
Detecting of eating from wristwork accelerometer	Smartwatch	Thomaz, Abowd [IUI 2015]
Detection of craving from respiration and ECG	Autosense	Kumar, Ertin [Ubicomp 2016]
Markers to detect behavioral and environmental risk factors		
Saccades and pupil dilation	iShadow	Ganesan, Marlin [Mobicom 2015]
Eye closure and blinks	iShadow	Ganesan, Marlin [Under submission]
Advertisement exposure	Eyeglass	Rehg [Under submission]
Stress assessment	AutoSense	Kumar [Ubicomp 2015]
POI datasets for tobacco/food outlet in Chicago	GPS/GIS	Patrick, Raab
Markers to detect physical state		
Internal lung and heart motion markers	EasySense	Ertin [Under submission]
General computational methods to infer markers		
Structured prediction for segmentation	All	Marlin [ICML 2015]
ML methods for lab-to-field generalizability	ECG	Marlin, Ganesan [Ubicomp 2016]
Analyzing compressively sampled ECG data	Autosense	Ertin [BSN 2016]

5. Thrust 2: Markers to Predictors (Jim Rehg)

Objective: Thrust 2 is working on developing data analytic tools for mining and visualizing patterns in marker data, predicting risk for adverse outcomes and designing Just-in-Time Adaptive Interventions (JITAI). Has split into a group focusing on predictive analytics and one focused on participant engagement. Both work closely with Thrust 4, which is conducting the smoking, eating and congestive heart failure studies.

Year 2 activities: Initial deployment of the Discovery Dashboard, developed ways to identify episodes of high stress and predict likelihood of future stress events.

Challenge: Noise and variability in marker data. Missing data due to physical activity & other confounding factors that make signal uninterpretable. Also

New methods are being worked on to predict risk for adverse outcomes in real time and deploy personalized interventions to improve health outcomes.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

5.1 Stress Detection and Prediction

Stress density can be used as a predictor to drive interventions.

There are two challenges: Validation and deployment, and replication of the process for other markers.

What have we learned from this process of detection with cStress and how do we build on it?

During discussion, it was noted that more discussion between disciplines was needed to avoid misunderstandings of how the algorithms work and how to use the markers. A loss of data resulted in a stress identification being made after the stress event had ended, and this led to a clinical intervention that made no sense to the user. More cross talk helped to resolve understanding of how to use the markers.

It was noted that the team needs to find ways to make the process more systematic with input from all disciplines.

5.2 Discovery Dashboard (Polack, Chen, Kahng, Sarker, Hohman, Fang, deBarboro, al'Absi, Sharmin, Chau)

The Discovery Dashboard was available for demo at the meeting.

In Year 2, the team released a browser-based Visual Cortex prototype for exploring sensor streams and applied it to the Minnesota dataset, which was a big accomplishment for MD2K.

In Year 3, the team will use MD2K feedback to make iterative improvements and integrate the dashboard with Cerebral Cortex.

5.3 JITAIs (Susan Murphy)

Lots of progress made on JITAI development by Susan Murphy's team. The team assessed moderation of time-varying intervention options to decide content and timing of JITAIs. An algorithm was developed and tested on micro-randomized data from HeartSteps, and the resulting paper is under review by JASA.

The team also is working to develop a "bandit" algorithm (for personalizing decision rules as a user experiences intervention). A paper from a thesis will be submitted soon. Three other papers, "Optimizing mHealth Interventions with a Bandit "(Rabbi, et al) was submitted as a book chapter; "Design Lessons from a Micro-Randomized Pilot Study in Mobile Health" (Smith et al) and "From Ads to Interventions: Contextual Bandits in Mobile Health" (Tewari and Murphy) are both to be included in the mHealth book edited by Rehg, Murphy and Kumar.

Operationalizing and defining components of a JITAI for behavioral audience (Nahum-Shani, et al.) is to appear in *Annals of Behavioral Medicine*.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

5.4 Overall plans for Year 3

Discovery Dashboard Release to MD2K Team

- Refine interface, visualizations, and analytic capabilities within Thrust 2
- Release to MD2K investigators for use in analyzing study data

Prediction of Risk for Adverse Outcomes

- Refinement of predictors for risk of first lapse in smoking cessation to incorporate geoexposure from GPS as well as self-report and stress marker data
- Development of predictors for risk of rehospitalization in CHF and participant disengagement.
- Initial evaluation of latent variable models for temporal risk prediction.

Micro-randomized Trials and JITAIs

- Develop JITAI-based methods to support interventions for CHF and participant engagement.
- Improved methods to learn decision rules for effective timing and content of just-in-time adaptive intervention based on micro-randomized trial data.
- Conduct analyses of smoking cessation micro-randomized trial data to assess proximal effects and moderation and enable refinement of simulation models.

6. MD2K Computational Platforms & Architecture (Mani Srivastava)

Thrust 3 objectives: scalable analytics & distributed real-time processing for mHealth.

6.1 Sensor Data >>Markers>>Patterns>>Prediction>>Intervention pipeline

Challenges: Volume, Velocity, Variations, and Versatility in the presence of battery, privacy and security constraints.

Activities: mechanisms & architectures, open interfaces, software development

- 2 components - 1. High frequency data collection 2. Cerebral cortex
- Tim's software has seen a fair amount of adoption thus far
- Tyson's team has been working on advancements: "big-debug"- debugging interfaces, Titian- trace back queries and execution replay, Vega- query execution

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Future directions:

- Storage optimizing for high-frequency streaming.
- Inferencing.
- Changing availability of devices/communication links (e.g., user leaves phone behind).
- Machine learning federation.
- Complex scenarios/complex event processing.

6.2 Privacy:

The obvious things have been handled, such as encryption to protect data from unauthorized entities. But hard privacy/security problems remain -- finding novel ways of de-anonymizing information like battery state to derive location, information leakage between sensors, between applications, covert channels, zero-day exploits. Outsourcing of computation offers its own pros/cons. Redistribution of sensor data can raise privacy issues as well, some of which our fields don't fully understand.

Providing users with agency throughout the chain of data collection; providing transparency about what happens with data as it moves through the pipeline. Privacy component within mCerebrum is an upgrade from simply removing the sensors.

Future directions:

- Understanding context-aware situations in which users' privacy preferences are predicted ahead of time.
- How to proceed with sharing raw sensor data vs. inferences made (some of which may be privacy-revealing).
- mSieve: privacy-aware sharing of sensor data (user-defined sanitization of raw sensor data that allows whitelisting vs. blacklisting of inferences).
- Make raw data indistinguishable from reference data regarding blacklisted inferences, without doing harm to the whitelisted inferences.

Papers were presented at Ubicom 2016, VLDB 2015, ICSE 2016, SoCC 2016, FSE 2016, Sigmod 2016.

7. MD2K Software Platforms (Tim Hnat)

7.1 Key capabilities

- Wide variety of sensors
- High-frequency raw sensor data collection
- Real-time quality assessment.
- Real-time derivation of biomarkers on the phone
- Sensor-triggered JITAI

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Allows for collection of raw sensor data up front and allows retroactive analysis.

Data arrives at phone through wireless interface. Sent to models. Used for engagement and then stored.

The architecture is multi-layered, using multiple individual applications. The goal is to reuse code to decrease development and processing time (example: puffMarker reuses respiration code)

Integrating algorithms into the software to allow information to stream now takes less time than it did at the beginning.

Are much better at integrating new sensors. Time frame is decreasing.

7.2 DataKit and DataKit API

- Public bus transports 3 million data samples per hour (75 million per day).
- It is extensible: publish/subscribe architecture to support data and computation reuse.
- Data processing is high frequency: Computation reuse through data streams, batch operations for system efficiency and a hybrid data store
- Support for privacy control is built in.
- Supports data offloading and archiving.

7.3 Real-time marker computation

Computation must not fall behind 3 million samples/hour. Can't run at the speeds that batch processing allows, but can compute every 60 seconds

Assess data quality every 3 seconds

7.4 Cerebral Cortex

- Big data companion to mCerebrum
- Built to support thousands of concurrent mCerebrum instances
- Currently powering 3 study types in 6 unique sites.
- 1,300 participants expected
- 50,000 participant days
- 144TB of raw data

It is currently hosted on campus, but will be ported to Amazon.

The primary focus at this time is getting the studies up and running. We also are looking at ways to be able to update software without impacting ongoing studies.

Challenge of the phone is getting data off. Pull data from phone every 15 minutes. Pull all raw data once an hour.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Everything that is done on the phone has to be done on the phone in terms of study requirements. (Info needs to be processed immediately to render JITAI).

Will have different versions in the cloud that will look at larger time ranges of data.

8. MD2K Standardization Activities (Ida Sim)

Standardization: how do we share everything out with common understanding?

Sharing biomarkers from multiple sources, with data coming from all different sensors and combined to be called stress, it is important to understand what stress is: data from sensors> data is processed> variables are calculated> share instance of stress.

Device agnostic data integration and FAIR data discovery for sharing are two key components of standardization.

The problem: Data Integration, Data Sharing

- Harmonize data collected from diverse devices
- Include metadata.
- Share processing algorithms.
- Expose measures from algorithms for external re-use.

The decision was made this year that data standardization from open mHealth does not belong on mCerebrum, but we do integrate data from various sensors.

Open mHealth - 72 publishes schemas with 10s more on the docket (example schema- probability of stress is a number between 0 and 1 that is probability of stress at any given time, determining unknown due to signal loss, determining “bias” and “windows”)

Metadata issues:

- Data stream vs. data point.
- What technical and clinical metadata is needed and for what?
- Which data points in which windows?
- Provenance pointer or token in Header schema.

Year 2 accomplishments:

- Schemas
    - promoted and sustained discussion on data sharing over Cerebral Cortex and related schema needs (e.g., p(stress)),identifying metadata requirements
    - Refined geoposition, survey (for Ecological Momentary Assessments) and medication adherence schemas for Year 2 studies.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

  - Refined physical activity and sleep duration schemas (for use in the Genome & Activity project)many schemas developed, defined schemas for how to share data based on "FAIR" sharing principles (Findable, Accessible, Interoperable and Reusable);
- Shims
  - Developed shim for MS Band (being discontinued)
- OmH-Md2K Data Integration Infrastructure
  - Focused on defining schemas and which APIs to use where
  - Pivoted from being driven by data integration needs to data sharing needs

Decision point this year: Do we share raw data or computer-derived data or real-time streams? Not sharing real-time streams, defer sharing raw data from cloud until the cloud is ready, focus on sharing biomarkers that are already computed (both internal like discovery dashboard, and external like Count Everything with 3 other BD2K centers)

Driving use cases:

- Is it useful to share when smoking vs. when stressed (defined: physical stress arousal) Eating detection?
- What is the purpose of the sharing? Mining for useful biomarkers? Placement of data in an EHR? Use of the data for research? To whom are data shared and why?
- What are data variables that we want to share? What is the metadata that we will share?
- It is critical to continue discussion of data sharing and use cases to narrow down our focus\
- We are currently sharing external data through Count Everything but there is a lot more work to do there

Year 3 plans

- Data Sharing: continue to clarify use cases,
- Architecture: Integrate Open mHealth Data integration with MD2K Asterix when available; work out where on-demand computation sits.
- Schemas: develop more MD2K-related schemas.
- API specs: upgrade with more filtering, meet Count Everything specs, handle authentication (Auth0), meet other external data sharing use cases.
- mProv: Implement provenance schemas and APIs
- Software: Provide downloadable instances of the MD2K-OmH platform

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

9. Sensor/Software Development Process (Tim Hnat, Emre Ertin)

9.1 Sensor Development Process: The Five Stages of Grief

- Denial
- Bargaining
    - Search in the commercial space. "This is a great band, but it won't work (aesthetics, battery life). "
- Eventually, acceptance for what will work:
    - Develop "laptop-size" prototype, followed by concept design stage and mock-ups to improve on previous sensor iterations.
    - Design and layout circuit designs to fit components for both function and form.
    - Find parts that can support the design, then develop firmware that can communicate with the devices, real-time OS.
- Test, validate, refine/finalize firmware.
- Ship to software team for integration.
- Ensure device feasibility (heat control; safety).
- Protection both on the battery and on the circuit to ensure shutdown if temperature exceeds 54 C.
- Discovery of process that damaged gyroscopes.
- Now: Have completely flexible architecture with new ideas on compressive sampling, wakeup on motion or context (e.g., use of Oral-B toothbrush for ROBAS work), agile platform to adopt new chipsets, next steps: multi-LED diode heart rate sensing.

9.2 Software Development Timeline

- Jan 2015: Tim joined full time
- Jan-May 2015: Cerebral Cortex prototypes and testing
- Mar 2015: mCerebrum development started (Monowar joined)
- July 2015: Initial Smoking Study design deadline
- July-Dec 2015: cStress (StreamProcessor) mobile implementation
- Aug 2015: mCerebrum Prototype (DataKit, AutoSense, PhoneSensor, Study UI)
- Feb 2016: MoodSurfing and Thought Shakeup completed
- Feb 2016: Last major modification to Smoking Study design
- Mar 2016: Nazir joined
- Mar-July 2016: puffMarker implemented
- Aug 2016: MotionSense and Omron integration completed
- Sept 2016: CHF design documents completed

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

9.3 Where we stand in the project and the progress that has been made

- For CHF -- OSU group has conducted pilot study to refine EasySense to ensure good quality data from the field study.
- Smoking -- iterative process including development of study design.
    - About 6-8 months ago, realized that no commercial sensors would fit the bill to detect smoking adequately for the entire day.
    - Emre turned attention to new wrist sensor that could fit study requirements and be sufficiently aesthetically/ergonomically appealing.
    - Most of us don't really understand what goes into the process of making a usable sensor.
- Software development has many moving components with considerable real-time constraints under which to operate.
- Sensor manufacturers have considerably greater human and capital resources.
    - An amazing effort was required by MD2K collaborators to deliver sensor hardware with integration into study software.
- Currently testing: Memphis, Northwestern, Ohio State.
- Sites using mCerebrum: Vermont, NW, OSU, Rice, UCLA, Memphis.
- There have been multiple sensor integrations (commercial and OSU-built).

9.4 Development process:

- Study planning
- Feature requests/triage
- Release planning.
- Issue/bug tracking.
- Software feedback/approval.
- Reiteration.
- 912 issues reported since May 2015
    - 212 bugs
    - 463 feature request
    - 117 tasks (e.g., change documentation).
- 716 issues resolved (over past ~350 work days -> avg. of 2/day)
- Comparison: Android has 12M lines of code.
- Research kit has 100k lines of codes amongst 84 contributors and 3 releases.
- ResearchStack has 17k lines of code across 11 contributors and 2 releases since Oct 2015.
- mCerebrum has 141k lines of code, 3 fulltime + 2 contributors in past 16-18 months.
- Nearly 1,700 commits and 410 software releases.
- Current status: Final phases of testing. All studies expected to begin in the next month. Modularity and configurability allows for greater flexibility moving forward.

9.5 Current Status

- First version is in final phases of testing

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

- All studies expected to begin in the next month
- Modularity and configurability enables:
  - New sensor integration time reduction
  - New biomarker models from collected data
  - New studies

9.6 Long term goals/sustainability of mCerebrum?

What is the long term goal/sustainability of mCerebrum especially within the context of other (some commercial) solutions?

One analogy is the TinyOS world that began in the research world and then spread amongst CS faculty. It grew with community involvement. The same paradigm would build mCerebrum.

The uniqueness of mC provides leverage over other options like ResearchKit (high frequency raw data rates is the key difference -- real-time computation of biomarkers from these high frequency raw data provides a constraint).

Other providers (e.g., Apple) don't have the market for raw data collection. The ability to make decisions in real time is important for intervention-dependent trials. Being able to make decisions quickly avoid wasted time pushing interventions that don't work.

10. MD2K Smoking Study (Bonnie Spring)

- Smoking and overeating tied for lead preventable cause of premature death
- "We have a wonderful study problem": determining whether smoking or overeating is the major cause, participants are embarrassed to report both, so we need an objective measure;
- These two things are inversely related in that weight gain can trigger smoking relapse and smoking cessation can trigger overeating
- Smoking and eating studies were initially separate and have been merged; interventions for both problems may be the same intervention
- Smoking, overeating, alcohol use trigger the same reward centers in brain.

10.1 MD2K Optimization Approach

- Detection algorithm
  - Objective, accurate, low burden, passive sensing of risk behaviors and trigger. Stress>>Smoking/Overeating.
- Predictive Model
  - Groundwork for dynamic, adaptive, real-time EMI.
- Adapt/Intervene
  - Derive decision rule for EMI from micro-randomized experiment.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

10.2 Study:

- 14 day protocol - 3.5 days participants are still smoking, this is to develop detection algorithms and to have adequate time for training;
- Self-reporting/interventions/EMAs (Event-contingent, Random, End Day)
- 3 Study aims: Detection algorithms (validate puffMarker; develop eating/overeating algorithms using wristband); Predictive modeling of smoking relapse; micro-randomized stress intervention during smoking cessation period of 10 days.
- Using 2 wrist sensors (MotionSense & MS Band), AutoSense, mCerebrum, GoPro cameras (field validation of smoking puffs and eating bites)
- Participants are given help to quit smoking. Trained on relaxation apps and skills
- Monitored 4 days pre- and 10 days post quit
- EM Intervention
- Interventions are prompted on MSBand and phone

3 custom stress regulation apps (locally designed) and 1 commercial

- Mood Surfing
    - Grounded in ACT
    - Target cognitive diffusion
    - Experts: K. Witkiewitz, I. Yovel.
    - Literacy level editor: A. Applegate
    - HCI: M. Sharmin; Programmer: M. Hossain
- Thought Shakeup
    - Grounded in CBT
    - Target cognitive restructuring
    - Expert: I. Yovel.
    - Literacy level editor: A. Applegate
    - HCI: M. Sharmin; Programmer: M. Hossain
- Head Space (Commercially Available)
    - Grounded in ACT
    - Meditation/Mindfulness
    - Consistently rated as one of the best 5 commercial meditation apps
    - Permission to free use in the trial
- Joy (Under development)
    - Story board phase
    - Grounded in positive psychology
    - Target relaxation + gratitude
    - Deep breathing exercise

Participant engagement is a concern, but participants are well-incentivized for wearing the sensors and responding to EMAs.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Because most smokers are unsuccessful in quitting - 93% fail in first week, failure is especially likely after one lapse. The goal of this study is to deal with this problem by intervening in real-time

If would be a big help if could know when they are about to slip. Study will answer whether stress is a useful tailoring variable as to when to deliver an intervention.

Stress is not a state, it waxes and wanes - we want to give smokers an alternative way to tailor to stress other than smoking. Participants will be trained to perform exercises in-lab so that they can be prompted in the field/real world.

10.3 Two Alternative Hypotheses: Using stress as a state of greater risk of relapse vs. using no-stress as a state of opportunity to improve resiliency to later stress (stress as a limiter of cognitive capacity)

Better time to prompt might be when they are not under stress. We will attempt to answer this question of whether stress is a good tailoring variable by delivering intervention or no intervention during both times of stress and non-stress

This study is novel in the clinical trials world because it is scalable (if it's not overly burdensome). Build best one-size-fits-all algorithm then tailor it as we go along.

How do you tell if an intervention is working? Have proximal (short-term) and distal (ultimate) outcomes.

Question is when to intervene. When the person is not stressed or when they are stressed.

Randomization will not occur when participants are unavailable: Can't provide a stress classification (bad data, physically active, etc.), person is driving (obviating safety concerns), or <60 minutes since last intervention or <10 minutes since last EMA (obviating burden)

Micro randomization probabilities set to yield avg. of 3 interventions/day across a.m., afternoon and evening w/50% under stress, 50% under no stress.

10.4 Year 3 plans

- Get all sensors working with mCerebrum.
- Enroll and complete data collection of version 1.
- Design and pilot version 2, which incorporates signal-contingent EMAs and refined microrandomization.

10.5 Expected outcomes:

- Detection algorithms (development of eating and refining of smoking).
- Establishment of predictive model of smoking lapse/relapse.
- EMI decision rule that optimizes when a stress regulation prompt should be delivered.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

11. MD2K CHF Study (William Abraham and Ramesh Emani)

11.1 Year 1 Protocol (pilot) Experiences:

Clinical goal: Develop a heart failure (HF) monitoring/management tool that can be used to reduce HF readmission rates.

Methods: Integrate the monitoring of clinical signs of congestion, other physiologic markers, and behavioral signs.

Importance: nearly 30% of patients with HF will be readmitted; major burden on patient and health systems. Disease is progressive - health often does not fully rebound.

HF follows something of a linear progressive course (sequential appearance of clinical signs: hemodynamic pressure changes, autonomic adaptation, thoracic impedance changes, and weight changes/perceived symptoms) that predate hospitalization

There are many points along the way where we are going to try to identify where to intervene.

11.2 Year 1 Recap

Developed novel sensor technology (EasySense) that can track the sensor's responses to clinical changes during HF hospitalization.

Patients with worsening HF identified early in their hospital stay; data acquired throughout the hospital stay.

11.3 Challenges:

- Heterogeneity
- Massive variations in underlying issues and causes.
- Concurrent diseases (heart attacks, irregular rhythms, etc).
- BMI variance.
- Patient response to therapies differs.
- Variable contraction of fluid spaces makes it difficult to determine where to measure fluid.
- Equipment design.
- Patient willingness to cooperate.

Between-patient variability is the thematic challenge with this HF population.

17 patients enrolled in the CHF pilot, with a low turn-down/declination rate (<10%) among patients. No specific negative feedback; improved feedback with updates in data acquisition protocol (e.g., fewer positional changes)

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

11.4 Year 1 progress: extensive troubleshooting of technology; after first 12 patients, we are getting much more reliable data

Phase shifting due to operating temperature of EasySense (resolved after first 8-10 patients); significant signal attenuation in body tissue

Patients 13-15 with more reliable, consistent data acquisition. Signal analysis in progress for patient 16, patient 17 just enrolled.

Overarching question: Over what time do changes become significant (is the time course days vs. hours?)

Importance: How frequently does data collection need to occur? (Is daily necessary -- probably from a scientific standpoint)

Goals for this year: Will consider signal amplifiers to improve quality; will transition to field study with moderate confidence that we can move data collection to outside the hospital

Patients have a good participation rate to daily measurements as long as the importance is explained and understood.

Importance is not just in the magnitude of volume, but location of the volume. Will EasySense detect smaller/incremental increase in fluid retention, especially prior to onset of symptoms? Other clinical signs (e.g., activity levels, fatigue) can also be important (aggregate clinical score may be the most useful).

11.5 Year 2 Field Study

Purpose: To evaluate the efficacy of the EasySense system in the field assessment of pulmonary congestion via thoracic impedance and cardiac/lung motion among patients with CHF. Looking at the 30-day post-discharge period (roughly 25% of CHF patients are readmitted during this time).

Study is about one year behind schedule due to vast amount of effort to software/study development.

Several mobile technologies employed: MotionSense, EasySense, MS Band, Omron BP Machine and Scale, self-report on study smartphones via mCerebrum (medication adherence), surveys of quality of life.

Prospective cohort of 75 patients. Inpatient phase and outpatient phase (30 days post discharge)

IRB submission to OSU on February 29, 2016; approval obtained on June 8, 2016

Enrollment slated for October 10, 2016.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Goal: learn what is important and apply it in an interventional study in years 3 & 4

Is a unique opportunity to see which of the sensors work best, alone or in conjunction with another. Ease of use will be an important outcome metric as well (which of these sensors is most informative as well as user friendly?)

We are confident we will find 75 eager participants for this study, but it is unlikely patients will accept the burden of this study in real-world applications

Year 2 study will be focused on collecting data, not acting on the results

Hopeful to recruit some patients with CardioMEMS to correlate pulmonary pressure changes with EasySense measurements.

12. Brainstorming & Publication Plans

12.1 Path Forward

Thinking about some path forward to bring in the JITAI component for future iteration(s) of CHF studies.

Can we derive a measure or combination of measures from this sensor data that are predictive and actionable? For example, there are predictive measures from implantable devices (high sensitivity), but are not very actionable (insufficient PPV). Also, what model for delivery of JITAI? Physician-directed patient self-care model might be useful. Incorporation of more behavioral interventions? For example, it may make be more helpful for patients to eat less sodium rather than increase diuretic dosage. No current sensors exist that investigate sodium homeostasis that may be influencing fluid retention.

Patients not taking meds is a major problem, studies from ~10 years ago show non-compliance in meds and diet.

JITAI could be a natural method if there is enough confidence in pushing the message (e.g., medication adherence) directly. Reliance on and craving for processed and fast foods is a major component of sodium intake. Behavior change focused on switching to healthier options like fruits and vegetables from unhealthier options could be helpful. The main point is reducing sodium intake, not necessarily weight loss. Cornell Tech has a new urine dipstick that determines the 24-hour sodium intake among patients with controlled hypertension, but it hasn't been validated in this population.

In order to make this work, we must be aware of the burden of too much information for clinicians. Best models are "management-by-exclusion with alerts" and "physician-directed patient-self-management."

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Another idea under consideration is to use geolocation in JITAI -- geofencing of grocery stores that reiterates reminders for nutrition best practices or even specific goals that could be customized based on the patient's needs/dietary plan. (This augments previously discussed JITAIs for fast food locations).

Bariatric patients have to have behavioral counseling before and after the surgery. Existing research on this might provide a knowledge base.

12.2 MD2K Participant Engagement Efforts & Plans

All of these studies rely on participants' compliance and long-term participation. Currently, we pay for their compliance through incentives.

Questions

- Are they truly engaged? We need people to follow the interventions and interact with the system in certain ways.
- How can we focus on their engagement in a systematic way?
- Using analytics to predict high risk for disengagement?
- Is there an explicit engagement component that we can build in?
- Are you doing them a favor with passive sensing, or are you taking away the cognition that is inherent to internal tracking of health? You want mindfulness (is cognition the precursor to mindfulness?). A semi-automated solution is ideal.
- We are designing an object to do magic for people and we don't have a clue how much magic is good for them.
- What encourages users/participants to engage with the phone/device? Currently, we interrupt them with prompts tied to incentives. Is there a universal answer? (Probably not)

For some people, writing everything down and tracking calories is helpful. For other people, it's not. Visual rewards may be necessary to keep participants engaged -- something of value.

Engagement for engagement's sake is not the point. Deeper engagement is -- surrounding activities that are beneficial (i.e., protective) for health outcomes. If you're doing it on your own, maybe you don't need the technology piece. Otherwise, maybe the system should revert to passive sensing until you need deeper engagement. Don't just assume we need high levels of engagement the entire way. Important to not just predict disengagement, but predict the need for engagement?

Interested parties can contact Bonnie to add an agenda item (or add it to the Thrust 4: Smoking Google Doc agenda) to continue engaging with this topic.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

13. Training Core

13.1 The Training Core continues to be a productive component of the MD2K Center.

- Workshop on software training day was presented yesterday.
- Online training resources.
- mHealthHub.
- mHealth Institute.

13.2 Year 2 Accomplishments

- Software development (mCerebrum, Software Development)
- Software distributions (public & private Github repositories)
- Training materials (mCerebrum, Cerebral Cortex) -- videos, print, web.
- In-app documentation and materials

13.3 Year 3 Plans

- Release end-to-end MD2K software updates.
- Support 6+ field studies running MD2K software for an expected 50,000 person-days.
- Transition private cloud to AWS.
- Continue support with training materials/documentation.
- Activate and moderate the online discussion forums.

13.4 mHealthHUB

- Web-based resources for training.
- Comprises student cohort, virtual student tutorial seminars (allied mHealth work).
- Hosts of regular webinars.
- Resources are archived on mHealthHUB (e.g., >35 hours of video which have >4,000 views).
- Webinars held monthly that include Q&A, archived on MD2K YouTube channel (e.g., ResearchStack, continuous eye monitoring, interactive computing).
- 48 articles
- 45 videos
- 15 MD2K webinars
- 18 other webinars
- Almost 16k page views in the past year
- New discussion tool, Discuss@MD2K, launched this week.

13.5 Year 3 Plans

- Continue student seminars

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

- continue webinars (Core is always looking to identify MD2K themes and recruit speakers, increase participation from other BD2K speakers, and develop curriculum models for dissemination);
- Seeking content contributors for mHealthHUB
- Continue posting relevant news as time permits.
- Seek content curators for mHealthHUB

13.6 mHealth Summer Training Institute (mHealth Boot Camp)

- Intensive session to train data scientists to work collaboratively with mHealth projects.
- Faculty interested in mHealth & data science.
- Students are treated as equal members of the mHTI teams and are given a unique opportunity to contribute and advance their work.
- The "winning" team receives a monetary award from P&G
- mHTI uses blended learning approach and is a great opportunity for networking.
- Next mHTI: August 6-11, 2017.
- Evaluative feedback from participants was very positive ("one of those seminal experiences in my career development")
- 88% of last year's cohort were junior or mid-level faculty; 67% female; 30% minority race
- Students report feeling they get a lot out of the program.
- Students to apply for 5 spots reserved for MD2K.

13.7 Year 3 Plans

- 5 students' trip to mHTI paid from MD2K Cohort
- Complete evaluation of program effectiveness and refine the program based on feedback.
- Increase efforts to recruit underrepresented minorities
- Integrate content from training institute to mHealthHUB so that researchers not selected for mHTI can also benefit, coordinate with BD2K's TCC training activities.

14. MD2K Future Funding/Sustainability Planning

MD2K is at its halfway mark. If we want to do anything to extend the grant, now is the time to do something.

We are still doing the work, so that its an indication that we like what we are doing.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Now that we have built this entire system, this center, along with the infrastructure to support it. Is it worth the effort to see if we can sustain this enterprise as a whole or just let it go when the funding runs out.

Abowd: The effort that was put into the software infrastructure would not have occurred w/out the center. I would be concerned about what would happen to it if the center went away.

Rehg: Relationships that we have built are the payoff.

Shetty: Second that. Initially met during gene and environment initiative. A lot of connections were lost when that ended.

Rehg: If people perceive that the connections are valuable and helping their work.

Kumar: The connection where we get to talk to each other and share experiences, is a lot more to it than funding.

Will begin exploring obtaining additional funding once I get a sense of the wishes of the room.

Kerr: The bar that NIH set for us was that we had to show that the technology was worth further investment.

Kumar: This center came together as a result of this grant. NIH would surely like to see this work survive and continue. Do you want and feel the need to continue.

Ganesan: This has been helpful in terms of exposing us to new problems and new ideas. This industry is not going away at any time soon.

Sensing will give way to the actuation side of things -- there are a number of ways to provide feedback re: wearables and their use in sensing and intervention.

Abraham: This has been very valuable. It's been incredibly difficult to build fully functional multi-disciplinary teams like this. The benefit will increase over time. It would be a shame if we got into a groove and had to end when the grant runs out.

How do we grow and shape this? Other areas where the tech we are developing can be applied to improve the human condition. Other successful companies/business require sustained effort after the team has been built -- which is very difficult in and of itself.

Kumar: There is a lot of talent in the room. The caliber of the team in the room will make obtaining additional funding not exceedingly difficult.

Abowd: Going forward do we want to stick with CHF and Smoking. Add to them?

Kumar: Several options we have: Let's hear from others first.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Marlin: The pipeline from circuits to clinical deployment is valuable. We'll be able to contribute more to the upper ends of the pipeline as more substantive data is obtained, but would need more time to continue ramping up.

Ertin: Very valuable to get the different viewpoints (being an engineer): Valuable having other input from other fields. Being realistic, not sure we can keep this large, monolithic entity sustained without bringing in more people and growing organically. Not just about maintaining this, looking at the same problems with the same minds -- allow for organic changes to occur to the group.

Kumar: Once we understand what we value, we can maintain the most valuable components and try to align ourselves with what is needed to obtain the funding that is out there.

Abowd: Emre, you're a fantastic resource for MD2K, but how do we find more of you?

Spring: It has been tough. I feel that it's too soon to say. It has been frustrating. There are times that my gaps in understanding are so great when I think about overcoming them. I feel removed -- if it worked, I'd be over the moon. Feel like we spent the first two years doing something wrong, and it's hard to feel optimistic. If things go well, then great -- it might be easier to feel optimistic then. It's indeed difficult to assemble such a diverse, talented interdisciplinary team.

al'Absi: We are going through team science. I have been in that zone before and sometimes still am. The tension can sometimes be good for creativity and problem solving. What to study if we go forward? Fundamental questions: Thinking about issues around physical activity & stress -- they apply to everything. If you can reliably assess and intervene effectively, you'll have profound impact on many conditions. Even going beyond the disease state to more fundamental questions.

Nahum-Shani: When you target mechanisms like stress, i think it's an opportunity to focus on the mechanisms we can sense but apply them across the board and bring in people associated with different health issues and work with them. Speaking to other health behaviors

Spring: Our contract was to work on smoking. Having spent so much time working on stress, that was one area that was challenging. I agree working on mechanisms, but we can't diffuse the focus too much.

Nahum-Shani: We have to show an impact.

Shetty: Plan A is that this whole thing is renewed. Plan B is see this as an infrastructure.

Infrastructure growing enterprise: We use our horizontal linkages to evolve our own communities, even via small grants. We have a backbone with the software to enable the community, and how do we create linkages amongst them.

Spring: We've devoted a lot of attention to student training. We also have some faculty cross-training to do. This is a frontier that we really need to make progress on. This is what will make a new trans-discipline to avoid working in silos and not being able to understand each other. We'll benefit the more

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

that we can understand what each other are doing. Not sure how to engage those mechanisms, but it's vital.

Shetty: Much of the student training is through the R25. Much faculty training occurs via the mHTI. Every time I finish an institute I'm wiser & more informed. I'm not sure where the funds and opportunities are to have that faculty-faculty conversation.

Kumar: This tension has been felt amongst many members of the team -- including the software team. Tim?

Hnat: I think having some sort of as an infrastructure and then have it expand in communities. If we can't have that pay off in 10 years, then what have we done? At some point, the software need to be the communities and that becomes the communities. Could MD2K become a node of a faculty summit (e.g., Microsoft's faculty summit) -- invite transdisciplinary people here and get them to talk.

Rehg: Broaden the participation?

Hnat: Yes, via training.

Rehg: Expanding outreach/marketing to expand the community.

Spring: Are there training mechanisms for contributing faculty?

Shetty: That was a separate OBSSR grant that came from a need to augment existing capacity. To follow up: The Stanford model follows an industry day to bring in faculty and exposed to each other and to industry. People who would productize/monetize.

Abowd: What can we accomplish in a one day gathering? High-level discussions only reveal our ignorance but do not find solutions. Example: Different disciplines have different understandings of stress. Using monthly calls to address these differences?

Nahum-Shani: The brainstorming sessions that Susan has been doing are focused, everyone gets together to understand what the idea is about and provide feedback while also correcting any misconceptions. Would this model be useful? People would be focused more on the problems than they are when just coming together under a grant proposal. The motivation would be learning, not just collaboration on a grant proposal. The PI would contact us with an idea -- can I come over? They would pay for the trip and spend 4-5 hours. Learn a lot about what each other does.

Spring: It's not a bad idea, but we would have to see a nucleus of expertise that this group had that is both unique and more expertised than others have. It might look like a day of consultation, and people here get built into grants for the sustainability aspect.

Nahum-Shani: Not everyone would find each session relevant: They choose to invest their resources or not. We can designate a certain amount of money to facilitate.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Hernandez: There are meetings for new collaboration initiative from NIH that can help facilitate new meetings for new avenues of science. I'm hearing this kind of thing might be eligible.

Spring: Start by looking at new collabs that involve people already engaged in this to enhance understanding and begin the faculty development that we need?

Shetty: We have a lot of resident wisdom in EAB member Dubinett -- any outsider's perspective?

Dubinett: This last idea about gathering around grants is an interesting one. We have an office that scours for RFAs and help teams to take care of the paperwork and barriers. One of the great choices in choosing your DBAs is that you've done this in a way that addresses the big transformation of what's going on in medicine. Applying all of the data to every individual patient is the biggest transformation in medicine since basing it on science. Our ability to utilize clinical data will happen by force by function of data warehousing.

More difficult thing is reaching out to activities of daily living and incorporating that into the universe of knowledge, and that is not a 5 or 10 year project, but it would be worthwhile.

Kumar: Sensors and software will evolve, but the approach is what we have created, and the perspective we have taken is how far we can bring the technology to help personalize each of these conditions, treatments, and can that move the needle. These are not new problems that we are addressing, but this team can try to see if we can make the next big jump in improvement. We can get up to 3 more years from the current mandate (with NCE). And look at other opportunities before the next two years are over to keep this momentum going. There will be other commercial sensors developed and things will continue to improve. This pipeline and iterative approach will help us learn a lot and find new improvements. Gradually we'll be on the cusp of a discovery to change medicine as we know it.

Rehg: Ideas from NSF side: Engineering Research Centers and Science and Technology Centers (ERC and STC Programs). We need to carefully think about the angle, but what we're doing fits these programs. Possibly a 10-year award. Strong ERC in mHealth? Casually, I think we should look into this very carefully. Next will come around in 2018. 1.5 year process. These are mechanisms we should consider.

Srivastava: ERC: needs strong industry support. Must become a priority for MD2K

Rehg: We can have that.

STC is more technology focused because of its nature, neither vehicle would be perfect, but It would be foolish not to explore both of these.

Abowd: Exploring clinical uses is critical -- feel we'd be a good fit for ERC.

Rehg: STC are broadening their scope

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Hernandez: STC is more flexible in looking at a problem. ERC is more engineering science focused.

Sim: app deadlines staggered so we could go for both.

Srivastava: ERC is more technology focused.

STC will probably come earlier than ERC deadline.

Rehg: we have some time (a couple of years). What is the portfolio of grants to bring in everyone and also broaden our team moving forward. Also the Expedition program from NSF, less money but still going strong.

Kumar: NIH Side: We haven't started any field study yet. Hopefully within the next month we can begin and having developed these tools and demonstrating their usability could trigger larger scale trials both for CHF and smoking/eating/stress intervention. We can think about follow up projects for the JITAI that are developed. The development of JITAI for other health conditions or fundamental health science. Several new biomarkers or predictors are being developed here but there are also many more to envision in the future. Remote studies, where we don't have to use second phone or chestband, where we send a smartwatch at most. What would that entail for tech development and that could be a re-iteration of MD2K (2.0). Other question: Our software, sensors are close -- but how much is MD2K able to help you with your other grants. What else should we be doing to make a stronger case when writing your other grants. It's fine to have an indirect role or direct role -- are there things we can do to offer more direct utility to your current or future projects.

Abowd: culling the RFAs and look on behalf of subsets within MD2K and identifying opportunities for new projects.

Kumar: one resource for that is Hernandez.

Dubinett: What's happened with NCATS/CTSAs they put large budgets into RFAs three times a year -- U01s that address infrastructure such as for a clinical problem $1m/year. CTSA institutions in this room could collaborate. They usually want 2 or more to collaborate.

Kumar: that will be nice as we advance the technology

Abraham: Logical next step after intervention study would be a multi-center RCT of 500-1000 patients. NHLBI multi center trials group has great interest in this arena. Need prelim data first.

Kumar: My hope is MD2K can provide sufficient memory data to help undertake the next step: results will go into clinical practice.

Some concrete ideas: Try and see how MD2K can help people in their projects that will come back to us to show what else we need in MD2K enterprise to scale and help as many people as possible.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Trying to look at various RFAs to see what we can facilitate in bringing new opportunities to people.

Quite willing and open to see how we can help to accelerate use of tech to advance health at individual level. Each step we take is closer to realize that vision. Other things to work out: With that kind of model, lots of modules with projects that are loosely tied to MD2K, still facilitating this kind of cross-talk across disciplines is quite valuable -- how to keep facilitating that. Brainstorming session is one idea, specific brainstorming on problems is another.

Another discussion tomorrow re: MD2K impact.

15. Consortium Core (Ida Sim)

15.1 Collaborative projects w/other centers:

- Disseminate MD2K Tools
  - USC (Enigma): brain imaging. Working with smoking cessation. MRI before and prior to quit, collect AutoSense data
  - Disseminates tools and methods, has funding, IRB approval, preliminary testing
- Convene BD2K around use cases
  - Stanford Mobilize: sensors that track activity and mobility.
  - Workshop March 2016 (invite-only). 22 high-profile companies (google, Apple, Samsung, etc.), 30 clinicians and researchers in a variety of disciplines
  - Jim Rehg participated in panel presentation
  - Next meeting 4/17-18/2017
- BD2K Consortium Working Groups
  - Doing work with validation working group (Catrine Tudor-Locke, Amherst), reviewing frameworks and methods and initiatives for validation
  - Use cases working group (Dr. Matt Smuck): Working w/particular diseases (i.e., Parkinson's). Tracking activity and mobility can be important for clinical approaches.
  - Goal is guidance for validation by industry and academia

15.2 mHealth Connect 2016

- Post-conference evals were overwhelmingly positive
- Data sharing group may happen next year (there was no one volunteering to lead it this year)
- Industry providing datasets to meet validation goals - MD2K may have a role
- Defines pre-competitive space with academia and science but opening a market space.
- 2017 meeting will be open to public for panel presentation first day; second day will be working meeting for invitees only.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

15.3 Count Everything Supplement

- 4 BD2K centers: MD2K, Santa Cruz, Harvard, UCSD BIOcaddie
- Goal: query distributed data sources to answer how many patients have a certain disease, genomic profile and mHealth data. Do it in a way that preserves patient privacy.
- Return counts across genomic, physical activity, and EHR data. Demonstrate query architecture across these discrete sources of the data.
- Each data source has its own API. EHR is Pic-Sure, Genomic is CBDTG, Activity is MD2K
- Supplement funding was received in Feb 2016

15.4 MD2K and CBDTG

- Combining genomic and mobile data around physical activity
- Goal: Demonstrate ability to query across real federated genomic and mHealth data
- ~200 people have put their genome data online
- Recruit from PGP
- Genomic data goes to Santa Cruz, health data goes to MD2K
- Queries around inflammatory genes and physical activity and sleep
- Will use one MS Band and 1 GeneActiv (research grade wrist sensor)
- Overall goal is demonstration/proof of concept for querying across these federated databases.
- Problem: Secure aggregating counts of relevant patients
- Synthetic data will be used to build the architecture and then be replaced by Genomes & Activity data.
- PIC-SURE: 40,000 Nhanes subjects
- CBDTG: 1000 subjects for genome project

Accomplishments

- Schema and global IDs
- Security Protocol
- Parsers

- User Interface
- Demonstration video shows how queries are made in Count Everything. Queries are fully customizable.
- Process has multiple steps. Synthetic data is available for query to test system.

15.5 Year 3 Plans

- Begin recruiting for Genomes & Activity project
- Maintain Count Everything project to resubmit for funding in Y4.

Clinical Informatics Working Group: disbanded this year

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Much of this work has been independent from MD2K, maybe we can find ways to collaborate/integrate

Important issue: even though participants give consent, do they really understand the implications of releasing all this genomic data?

Need to find out what people want and how open datasets need to be. Who would use it, why would they use it? Do we allow other people to donate datasets? What datasets would really help?

Jim R. People who study chronic conditions would probably love to get their hands on data to determine how it relates to disease.

Can generalize location data to home, work, a restaurant without specifics and it will still be useful.

Jacqueline K. At our work, we want to know where a study subject is and what is around them -- geoexposure is important. Our approach is to build secure spaces where researchers can work.

Ida: Do we allow datasets to be downloaded or do we provide a specific environment to work with datasets.

Vivek: when we are working with vulnerable populations, do they really understand what sharing their data means? Need to engage a bioethicist and organize a bioethical framework.

Shahin: Need to define what the research question is. A person consents for sharing the data for one question, then can it be shared for another question?

Santosh: Need to be able to expose a certain amount of metadata so that researchers can determine if they need it.

16. Data Description, Sharing, Access, and IRB

16.1 Year 3 datasets

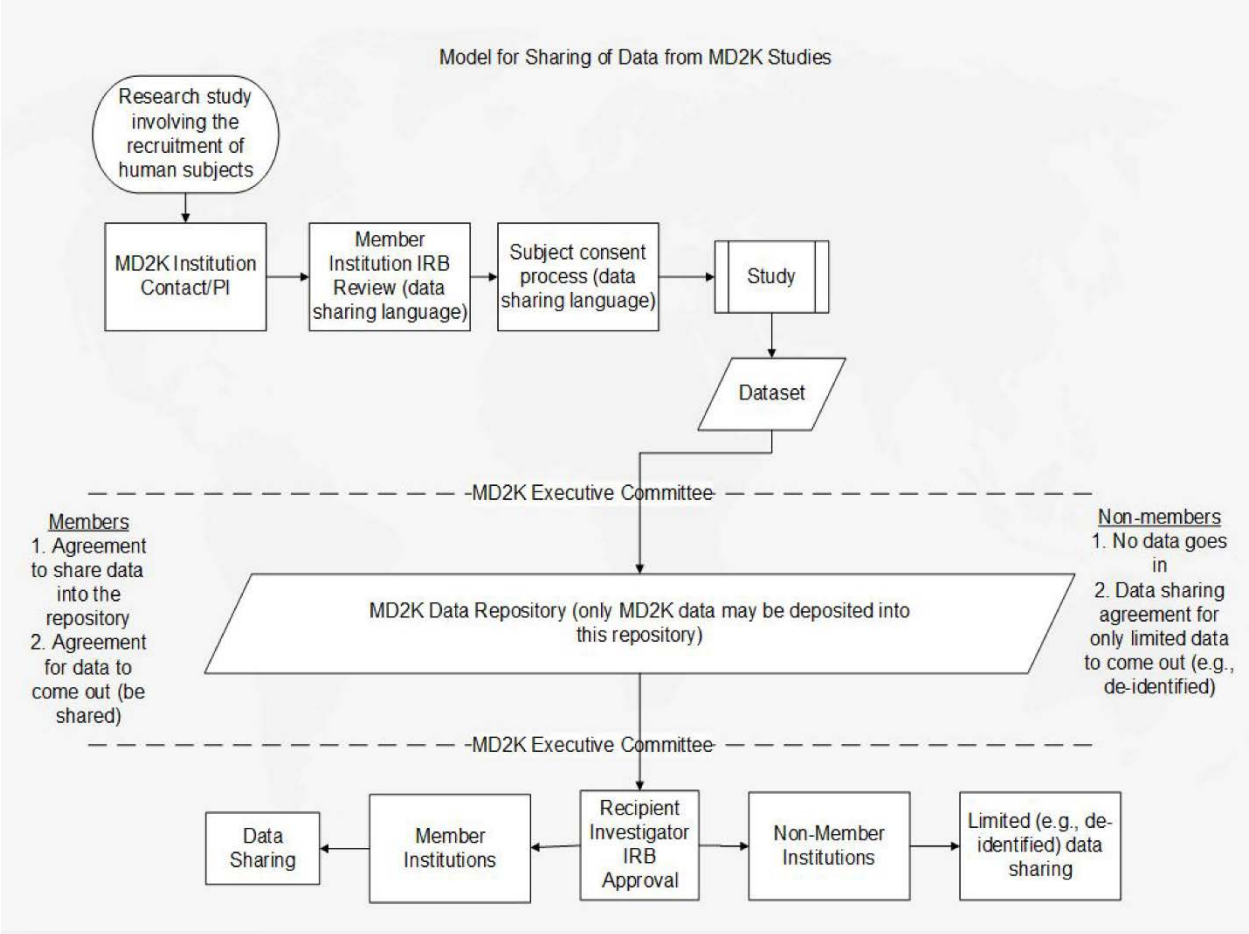
- Smoking study
- CHF study
- Human Genome project
- ENIGMA

16.2 Model for data sharing

- Begins with research question
- Consent
- Study of human subjects

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

- Data set
- Data repository
- Data sharing
- Only MD2K data goes into the repository.



16.3 Pre-MD2K Datasets

- Conversation Detection
- AutoSense Smoking Pilot
- Smoking Prediction
- Everyday Activity Recognition

16.4 Year 3 datasets using mCerebrum (not necessarily MD2K studies)

- Socioeconomic status, smoking cessation (Rice/Wetter)
- Smoking-related disparities among African American smokes (Rice/Wetter)

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

- Oral Behaviors (Memphis/Kumar)
- ROBAS (UCLA/Shetty)

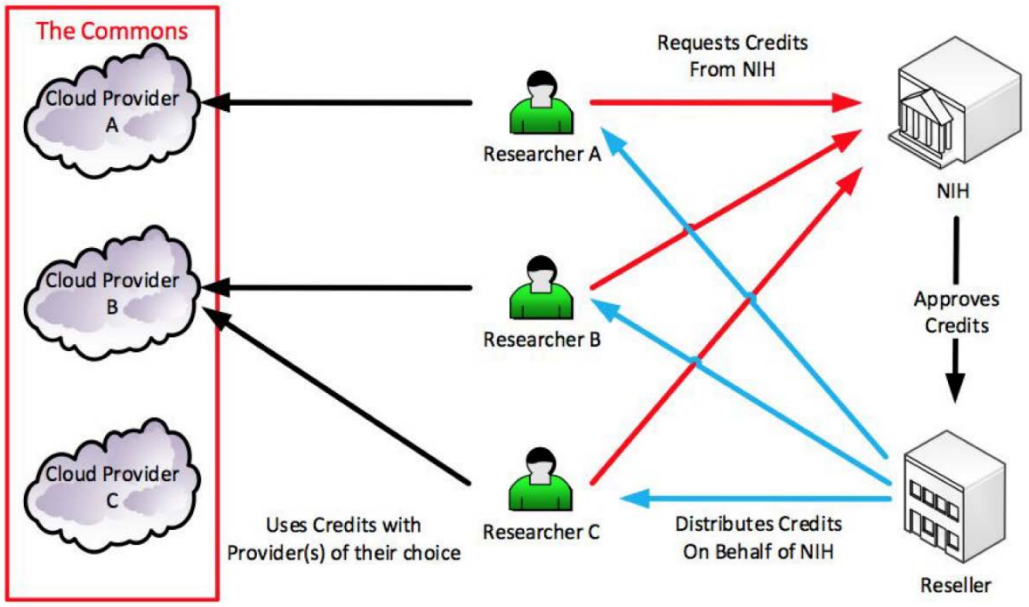
SocioEconomic Status, Stress & Smoking Cessation

- Most profound Health disparities cncentrated among lower SES
- Low SES leads to more stress which influence health behaviors
- Study aims to define links between SES, Stress and smoking lapse/abstinence; links between environment/context and biobehavioral/psychosocial predispositions with SES, smoking lapse/abstinence and stress.
- Examine relationship between stress and smoking lapse
- 300 smokers of all races who are all low SES across 26 weeks receive smoking cessation treatment (patch therapy, etc) using AutoSense for two weeks (similar to NWestern study but without intervention) as well as Motionsense wrist and up to 12 EMAS a day.

ROBAS

- Oral health behaviors
- Applying mHealth to oral health (important because of its relationship to poor overall health)
- Aims: 3 phases: start by suing simulations and test subjects then refine algorithms for fields studies and finally a field evaluations with 120 participants

16.5 New outlets for Big Data: NIH Data Commons (cloud credits), bioCADDIE (Data Discovery Index), etc.



MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

16.6 Sharability

- Cleaning
- Feature Extraction
- Meaningfulness/usability
- Compliance (IRB/Consent/Human Subjects Protections)
- Data sharing agreements (co-authorship expectations, third-party sharing)
- Currently working through issues systematically for MD2K. Issues are intrinsic to datasets and research of this type.

16.7 Next Steps

- sIRB (single-Institutional Review Board)
    - DHHS/NIH moving away from multiple IRB review
    - Single IRB reviews multi-site studies (But which??)
    - IRB Authorization Agreements (My IRB says that your IRB says it's OK)
    - What's the point? Streamline the process; maintain rights of subjects
    - Understanding this process for next study iterations

Research offices at universities across the country are currently working to understand the sIRB process, over the next year we should make a lot of progress in this area

Shetty: sIRB will be especially valuable for smaller/rural institutions that often do not have the infrastructure to handle IRBs

Co-authorship guidelines

Approved in September

Can be found here:
https://docs.google.com/document/d/1LPWd7MKv_ykSOBf11ucT5vZrNio6RhSTyl5AeSvrxJ4/edit

17. Administration Core (Joe Biggers)

17.1 Mission of Admin core: Tie everything back to the MD2K Mission Statement

17.2 5 components

- Resolving issues

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

- Applying the concepts of team science across geographies
- Achieving practical integration of core activities through input
- Managing resources, spreading meaningful information.
- Customer service based -- accommodate everyone to the best of our abilities.

20 faculty, 12 institutions, 6 disciplines, 4 postdocs, 18 students, 8 collaborators, 9 FT staff in Memphis (77 people) and growing

17.3 Y2 achievements

- Expansion of Operations staff from 6 to 8 – (business officer, technical writer, software engineer)
- Held last year's AM with 80 investigators/attendees
- 200+ research papers accepted for publication
- Almost 40 hours of training videos released
- 118 MD2K talks at 81 global meetings
- 61,925 page views at MD2K.org
- 140 Twitter followers '@MD2Korg'
- 38 inquiries received from over industry contacts
- 146 MD2K milestones completed so far
- 1658 GoogleDocs
- mHealth training offered to 70 students and 30

- faculty from 15+ institutions in multiple disciplines
- MD2K mentioned in 41 articles
- Mention in Science magazine with ~8,000 views
- Almost 252 man-hours on the telecons (avg. of 8 emails per call)
- Google named Memphis this digital capital of TN for 2015, citing MD2K as primary reason

17.4 Year 3 Plans

- Hold Y3 AM (Completed October 5)
- Hold first hands on software day (completed Monday October 3)
- Hold additional in person meetings for subgroups
- Year 3 - Transitioning to cloud via credits program, sustainability to be explored (create formalized plan)

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

18. MD2K Impact (Santosh Kumar)

How can we improve MD2K to impact and help more people?

Work products of MD2K

- Software
- Biomarkers
- Sensor-triggered intervention
- Training materials

How can MD2K be more useful to investigators?

Suggestions:

- Apply MD2K science to other projects
- Use training materials and student work as course modules
- Train investigators in other locations on how to make minor changes to apps/software
- Create a web-based configuration where PIs/coordinators can define their own study. Need software support to implement.
- Send Memphis team to other locations to train their software people
- Provide version of the software without all the bells and whistles that can be implemented quickly.
    - The mobile platform is highly modular. Each sensor is an independent application that talks to the DataKit app. You should just be able to pick and choose the apps you want. Every app has its own independent repository and releases on Github.

Also, how can MD2K help Health Investigators, data science investigators, students and instructors, industry, individuals, workplaces, healthcare.

Is there some way that MD2K can be put to more use in your other endeavors? Get to market faster, get a study faster

Abowd: Another project I'm working is trying to do large-scale exploration with mobile devices. I'm wondering if MD2K is a part of that solution. Also, think about what might be done with MD2K that can be useful in classrooms.

Kumar: Training materials and student work could be used as course modules

Spring: We'll probably want to pilot/tweak the study before running the full cohort of 75 participants. However, changing things requires modifying the technology. Could we do n of 1 pilots to tweak things rather than a complete separation of the tech in Memphis and the pilot happening elsewhere (e.g., NW).

It would be nice to have the ability to make changes.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Hnat: If it's configuration-based, we can give instructions/tutorial (easier). If it's application/development based, it can be more difficult without development/developer access.

Spring: One example… give participants information about when they were stressed.

The relaxation effect of MoodSurfing, for example, might be obviated a little by the end of the application, being able to make tweaks like those might be helpful.

Kumar: Excellent point. Can we imagine more adaptation of the software -- it is open source. What might be more easy adaptation, what are the more desired adaptations?

Hnat: We've always thought about having a web-based config where PIs/coordinators can define their own study. Need software support to implement.

Spring: This would be essential for uptake down the road.

Sim: That's a barrier. People could imagine using this, but then realize they don't have the skills to customize the code. They'd probably have to go to a company if they don't have access to a developer.

Hnat: What is easy-to-configure software from a clinician standpoint.

Sim: that will only get us so far, and it's probably not far enough (we have a range of use cases). Though it is a good start. The brokering function between clinician/developer is critical. "This is how you talk to a tech person" "Here is a sample contract" etc.

Kumar: We could try to train some of Bonnie's developers.

Spring: This may be unrealistic but even to aim for a degree a modularization that could be done with just the smoking app.

Hnat: Some level of modularity exists within the configuration files. Almost every feature seen in the platform can be changed via configuration (Kumar: it's like a script). The UI powers all of the studies with different config files.

Spring: I employ 3 programmers, so we may have a good test case. Maybe we should try it.

Kumar: Maybe we can have another software session with these programmers and see what can happen of it.

Spring: Would love for you to come to NW. Tech works so much better in Memphis. Think about software training day in NW?

Abowd: There is a growing community of teams trying to provide solutions similar to MD2k. The way people make money out of this is by becoming consultants.

Take the open source and configure it into what's usable.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Kumar: This could be a good model (Hnat: would help with dissemination, too)

Abowd: Rehg and I hired Prometheus to give people remote access to our data. There are plenty of examples of companies who do this.

Kumar: We can begin by trying to train Bonnie's developers.

Spring: We have been doing interdisciplinary science here, not transdisciplinary. Everyone working separately-but-together. Colocate virtually/mentally to reduce the limitations in moving forward.

Carini: The experience of the first two studies has shown us all the complexity, but also what we can do preemptively before future studies. Dashboard, for example, is a good idea so that people can start playing with things to know what questions to ask.

Kumar: It will be great to see if Abowd's project will adopt MD2K, and create more knowledge about who can use our software for the greater good.

Ertin: As the design of software was done by committee to add features, it became more complex. Would it be use to have a version of the app not at all configured that has all of the sensors enabled. Click "go" and it just starts collecting data with no bells/whistles, EMAs, identifiers. Students could just start collecting data without needing to learn of other features like EMA, etc.

Hnat: once we get studies rolling, this is a high priority task. To get generic way of collecting data. It is currently a manual process.

Ertin: I select the ones I want and click start/stop. Simple enough to give to an undergrad and say "see if you can detect X."

Hnat: This is what we want.

Spring: That may be helpful for what Abowd is trying to accomplish: improving student health across campus. This could be helpful for people working on Campus Life.

Kumar: Emre's suggestion could help with course projects for students.

Sim: Some will want to use their own or commercial sensors, perhaps MD2K sensors.

Hnat: We're targeting only raw data sources.

Sim: Not many people are only using raw data, especially on the health side -- not many have the capacity to understand the raw data (much less: how to get the raw data). How to

We should think about additionally supporting sensors that are more commonly used even if they don't provide raw data.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

David Conroy: Simulated data for students? That would be very helpful for their purposes.

Spring: We feel limited by the lack of available hardware suites.

Kumar: We can send more sensors home with you.

Chau: How modular is the back-end? Students sometimes take Github projects from here and there and see how to combine them for their purposes. Is it possible to just take the piece(s) they want.

Hnat: Highly modular for the mobile platform. Each sensor is an independent application that talks to the DataKit app. You should just be able to pick and choose the apps you want. Every app has its own independent repository and releases on Github.

Abowd: question about scientific impact: Is there an opportunity from our 2 studies to publish in the "best place you can publish"?

Spring: Yes, I think so.

Kumar: Strong potential for this.

At least some landmark publications describing the nature of the data collection. While we don't have the eating marker implemented on the phone yet, but it can be applied post-hoc after those algorithms have been refined. This capability can be highlighted once we have the data. Similarly, we don't yet have the real-time geoexposure system, but because we are collecting the raw data, we can go back and infer the geoexposures later on. Similar with refinements to existing markers like stress.

Planning a workshop: We have smoking, eating, CHF, Vivek is doing oral health monitoring, NIDA wanted to adopt the cocaine use detection from heart rate data. Various applications exemplified in this approach/with these methods. Other health targets? Should we think about convening a planning workshop to expand what is possible/what can be done with this data collection approach? Ex. autism work done in GA Tech (Abowd/Rehg).

Abowd: How does this dovetail with your co-location, Sim?

Sim: Makes sense to combine them. Pulmonary, mobility markers, a number of others (depression)

Kumar: Bringing together people from various domains to see what capabilities would be needed to be able to make this applicable to wide variety of health conditions.

Sim: What will attendees get out of it?

Kumar: Right. This approach of collecting raw sensor data -- does it facilitate new discoveries that were not possible before?

Abowd: Would they come if training on mCerebrum was included?

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.

Sim: I think they need to be convinced that there is value/ready for prime time, examples of how used, there is a way to collect the raw data -- it's so far away from what they are thinking about. Maybe it's initially a webinar/a series. This is how raw sensor data has been used for autism -- now we bring together people from pulmonary/NHLBI and have a deep dive as an initial foray into structuring a workshop outlining a series of grants to apply for. There has to be something to draw people. Initial explorations in high-value clinical domains?

Kumar: Have one focused on each specific domain?

Sim: Yes, pulmonary folks want to hear pulmonary.

Dubinett: Having people from federal funding agencies included in these meetings would be a great idea.

DoD, NIH, other institutes NCATS leadership would be good to have at the table to see how new trial designs could be implemented with this infrastructure. Helps to bring the investigators to the table when they recognize that funders are at the table -- thinking about the next phase.

Kumar: There are many possibilities.

Two things: Supply side, demand side. We want them both to converge. Supply: How to make the software easily usable, configurable, when people are interested they see how to use it. Demand: Create demand by letting them know this exists and what it can do -- now do you want to explore it?

Make efforts from both sides, but we better be ready to meet demands when they come. Research, industry, etc.

19. Conclusion (Santosh Kumar)

Kumar: Will poster presenters be willing to spend more time with their posters if possible -- ensure more people can see your work.

Carini: Monday meeting was really awesome -- whoever had the idea -- organization was great. Enjoyed it, presentations were great. Nice to get a sense of other things we don't know. To see the application, the sensors, it was great. Liked the presentations yesterday e.g., Emre -- the process of development. Tim's presentation -- getting a sense of what happens behind the scenes. We usually only see the end result. (This gives huge context)

Kumar: Glad you said that. Team has spent great effort -- glad it was well-received. Thanks to all for coming. Looking forward to talking again.

MD2K is supported by the National Institutes of Health through the NIH Big Data to Knowledge (BD2K) initiative.
For more information, please visit www.md2k.org.