



A Challenge in Developing Learning Algorithms to Personalize Treatment in Real Time

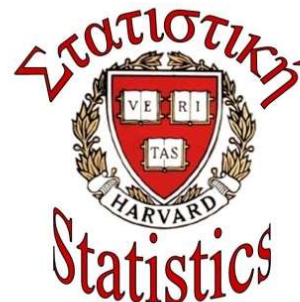


JOOL
HEALTH



BariFit

Susan A Murphy



The Methodology Center
advancing methods, improving health

Happy Birthday Aad!



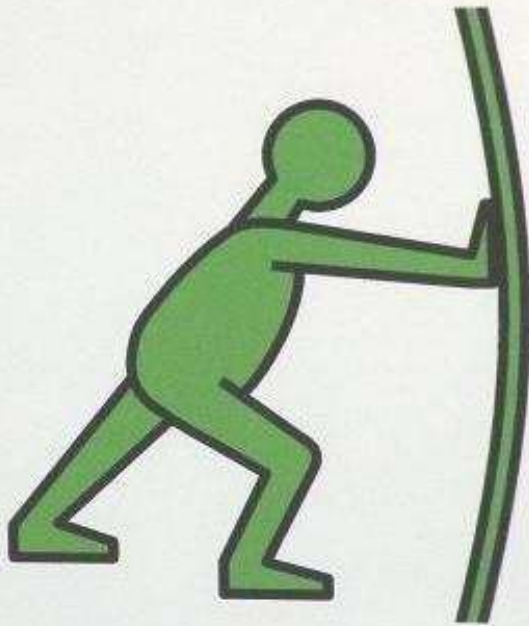
60 candles.....

Outline

- Review of Mobile Health Treatments & Reinforcement Learning
- A Challenge to Sequential Experimentation
- Feedback from You on a Potential Solution or Other Solutions (?!)

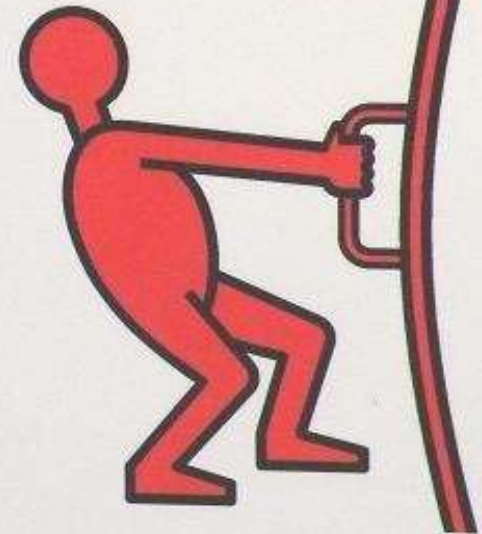
Mobile Health Treatments

P
U
S
H



push pull signs.com © (R)

P
U
L
L



push pull signs.com © (R)

Continual Learning

- Sequential experimentation using online forecasting and reinforcement learning algorithms
- Permit causal inference analyses at study end and/or interim monitoring and causal analyses when implemented by health system

Data

- On each user: $S_1 A_1 R_2, \dots, S_t, A_t, R_{t+1}, \dots$
- t : time
- S_t : Context accrued after $t-1$ and up to/including decision point t
- A_t : Action at t^{th} time (treatment)
- R_{t+1} : Reward (e.g., utility, cost) accrued after time t and prior to time $t+1$

Reinforcement Learning (RL)

- RL algorithms use sequential experimentation to select the action A_t after observing context, S_t . To learn about effects of actions,
 - Stochastic alg. use external variation, e.g. randomization
 - Non-stochastic alg. use internal variation
- Objective is to maximize total rewards or discounted total rewards: $\sum_{t=0}^{t=T} \gamma^t R_{t+1}$ over a horizon, T

γ is discount rate

Reinforcement Learning (RL)

- RL algorithms use sequential experimentation to select the action A_t after observing context, S_t . To learn about effects of actions,
 - Stochastic alg. use external variation, e.g. randomization
 - Non-stochastic alg. use internal variation
- Objective is to maximize total rewards or discounted total rewards: $\sum_{t=0}^{t=T} \gamma^t R_{t+1}$ over a horizon, T

γ is discount rate

Thompson-Sampling Bandit Algorithm

- Highly successful and widely used stochastic alg
- Algorithm:
 - 1) Bayesian regression model for $r(s, a) = E[R_{t+1} | S_t = s, A_t = a]$.
 - 2) In context, $S_t = s$, select binary treatment, $A_t = 1$, with posterior probability that $r(s, 1) > r(s, 0)$.
- Common use employs a parametric model for $r(s, a)$ (linear, neural net,...)

Mobile Health Challenges

- Variety of stakeholders with different inferential goals including causal inference.

A Sol'n: Randomization (via posterior distribution) + clipping to ensure non-zero randomization.

- High noise.

A Sol'n: Algorithm pools information across users in real time, online, to construct posterior distributions.

Today's Focus

- Variety of stakeholders with different inferential goals including causal inference.

A Sol'n: Randomization via posterior distribution + clipping to ensure non-zero randomization.

- High noise.

A Sol'n: Algorithm pools information across users in real time, online, to construct posterior distributions.

Toy Two-Arm Bandit

- Working model:

$$R_{t+1,i} = \beta_0 + (\beta_1 - \beta_0)A_{ti} + \varepsilon_{ti}$$

where ε_{ti} are iid $N(0, \sigma^2) \forall i, t$

- Prior: β_0, β_1 are iid $N(0, 1)$
- Time $t = 1$
 - $i = 1, \dots, n$ users move through study synchronously,
 - Treatment: A_{1i} are iid $Ber(\pi_1)$;

$$\pi_1 = P(\beta_1 - \beta_0 > 0) = \Phi(0) = 0.5$$

Toy Two-Arm Bandit

- $i = 1, \dots, n$ users move through study synchronously

- Time $t = 2$

- Treatment: A_{2i} are iid $Ber(\pi_2)$;

- $\pi_2 = P(\beta_1 - \beta_0 > 0 | H_1)$

- $H_1 = \{(A_{1i}, R_{2i}), i = 1, \dots, n\}$

- $\pi_2 = \Phi\left(Z_n + \sqrt{n} \frac{\beta_1 - \beta_0}{2\sigma}\right), Z_n \xrightarrow{d} N\left(0, 1 + \frac{(\beta_1 - \beta_0)^2}{4\sigma^2}\right)$

Randomization at Time $t = 2$

Treatments $A_{2i}, i = 1, \dots, n$ are independently randomized in $\{0,1\}$ with probability

- $\pi_2 = \Phi \left(Z_n + \sqrt{n} \frac{\beta_1 - \beta_0}{2\sigma} \right),$
 - $Z_n \xrightarrow{d} N \left(0, 1 + \frac{(\beta_1 - \beta_0)^2}{4\sigma^2} \right)$
- If there is no true difference between the arms, $\beta_1 - \beta_0 = 0$, the randomization probability π_2 does not converge to a constant

Small n and small $\frac{\beta_1 - \beta_0}{2\sigma} \rightarrow \text{instability}$

Data from Toy Example

- n users each with: $S_1, A_1, R_2, S_2, A_2, R_3, S_3$
- S_t : Context accrued after $t-1$ and up to/including decision point t
- A_t : Treatment at t^{th} time
- R_{t+1} : Response accrued after time t and prior to time $t+1$

Scientists will want to conduct a variety of analyses using this data

Instability

- In the trial, the distribution of the binary treatments $A_{2i}, i = 1, \dots, n$ are independently randomized with probability π_2 .
- For low signal, e.g. $\frac{\beta_1 - \beta_0}{2\sigma}$ small, π_2 varies across replications of the trial -- approximately Uniform.
- Usual large sample approximations are no longer valid when the A_{ti} 's form the design matrix or otherwise enter, after study, analyses of treatment effects

Consequence of Instability

Poor small sample normal approximation

- Well-known problem in the bandit literature.
 - Lai and Wei's (1982) treatment of OLS estimators includes a stability condition that is violated
 - Dvoretzky (1972) provides a toy example illustrating deviation from normality
 - Deshpande, et al., (2018) provide a correction for OLS estimators, valid under a weaker stability condition

Our Goals

- Use stochastic bandit or stochastic RL algorithms to personalize during trial; pool over users to speed up learning and relearning for each user.
- Facilitate a wide variety of after-trial causal inference analyses (different outcomes, classical longitudinal analyses,...)

Instability challenges these goals

Idea

- RL algorithms, that use the posterior distribution to conduct randomization, can be viewed as conducting an hypothesis test, $H_0: \beta_1 - \beta_0 = 0$, and then randomizing according to p-values.
- Yet the traditional choice of the prior gives 0 probability to $H_0: \beta_1 - \beta_0 = 0$.
- Bayesian thinking re hypothesis testing/model selection offers an alternative approach.

Idea: Toy Two-Arm Bandit

- Working model:

$$R_{t+1,i} = \beta_0 + (\beta_1 - \beta_0)uA_{ti} + \varepsilon_{ti}$$

where ε_{ti} are iid $N(0, \sigma^2) \forall i, t$

- Prior: u is $\text{Ber}(p)$ and β_0, β_1 are iid $N(0, 1)$
 - $i = 1, \dots, n$ users move through study synchronously,
To randomize A_{1i} draw $u \sim \text{Ber}(p)$
 - if $u = 0$, then draw $A_{1i} \sim \text{Ber}(\pi_0)$;
 - if $u = 1$, then draw $A_{1i} \sim \text{Ber}(P(\beta_1 - \beta_0 > 0))$

Toy Two-Arm Bandit

- Working model:

$$R_{t+1,i} = \beta_0 + (\beta_1 - \beta_0)uA_{ti} + \varepsilon_{ti}$$

where ε_{ti} are iid $N(0, \sigma^2) \forall i, t$

- Prior: u is $\text{Ber}(p_1)$ and β_0, β_1 are iid $N(0, 1)$
- Time $t = 1$ Treatments, A_{1i} , are iid $\text{Ber}(\pi_1)$;

$$\pi_1 = p_1 \Phi(0) + (1 - p_1)\pi_0$$

Toy Two-Arm Bandit

Time $t = 2$

For each user i , draw $u \sim \text{Ber}(p_2)$,

$$p_2 = P(u = 1 | H_1)$$

- if $u = 0$, then draw $A_{1i} \sim \text{Ber}(\pi_0)$;
- if $u = 1$, then draw $A_{1i} \sim \text{Ber}(P(\beta_1 - \beta_0 > 0 | H_1, u = 1))$

where

$$\bullet H_1 = \{(A_{1i}, R_{2i}), i = 1, \dots, n\}$$

Toy Two-Arm Bandit

Time $t = 2$

Treatments: A_{2i} are iid $Ber(\pi_2)$;

$$\pi_2 = p_2 P(\beta_1 - \beta_0 > 0 | H_1, u = 1) + (1 - p_2)\pi_0$$

where

- $p_2 = P(u = 1 | H_1)$
- $H_1 = \{(A_{1i}, R_{2i}), i = 1, \dots, n\}$

Randomization at Time $t = 2$

Treatments $A_{2i}, i = 1, \dots, n$ are independently randomized with probability

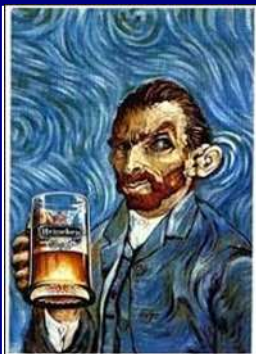
$$\pi_2 = p_2 \Phi \left(Z_n + \sqrt{n} \frac{\beta_1 - \beta_0}{2\sigma} \right) + (1 - p_2) \pi_0$$

- $Z_n \xrightarrow{d} N \left(0, 1 + \frac{(\beta_1 - \beta_0)^2}{4\sigma^2} \right)$
- $p_2 = P(u = 1 | H_1) \xrightarrow{p} 0$ if $\frac{\beta_1 - \beta_0}{2\sigma} = 0$

The randomization probability π_2 converges to a constant including when true difference between the arms is 0.

Thoughts/Questions

- This type of prior has not been used in Bandit or RL algorithms
- Have you seen Bayesian hypothesis testing/model selection ideas used in Bandit/RL/sequential experimentation?
- General Problem
 - The number of time points, T , is large
 - Nonstationary
 - Full RL with context

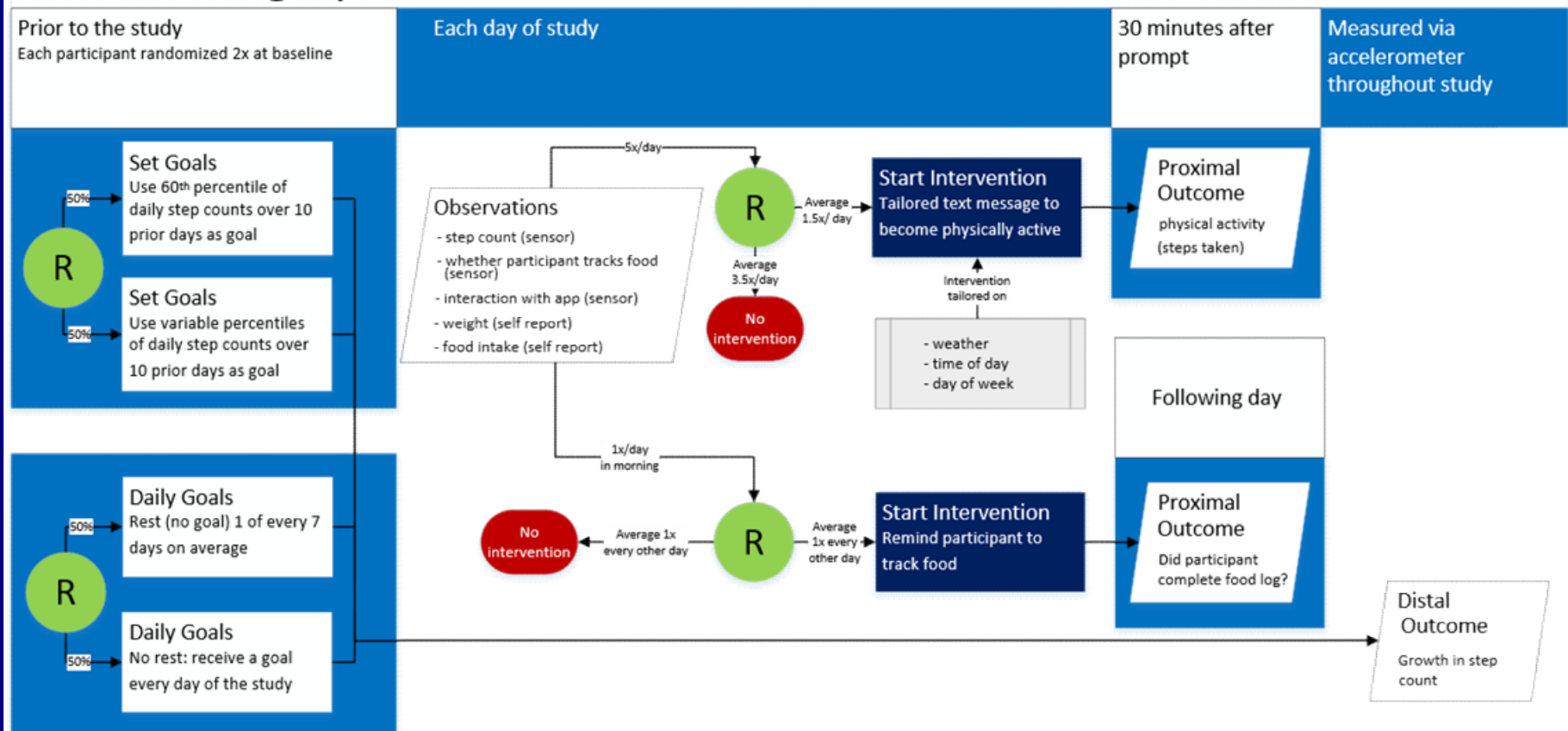


Happy Birthday Aad!



BariFit

BariFit MRT to Promote Weight Maintenance Among People Who Received Bariatric Surgery



PI: P Klasnja

Location & Funding: Kaiser Permanente

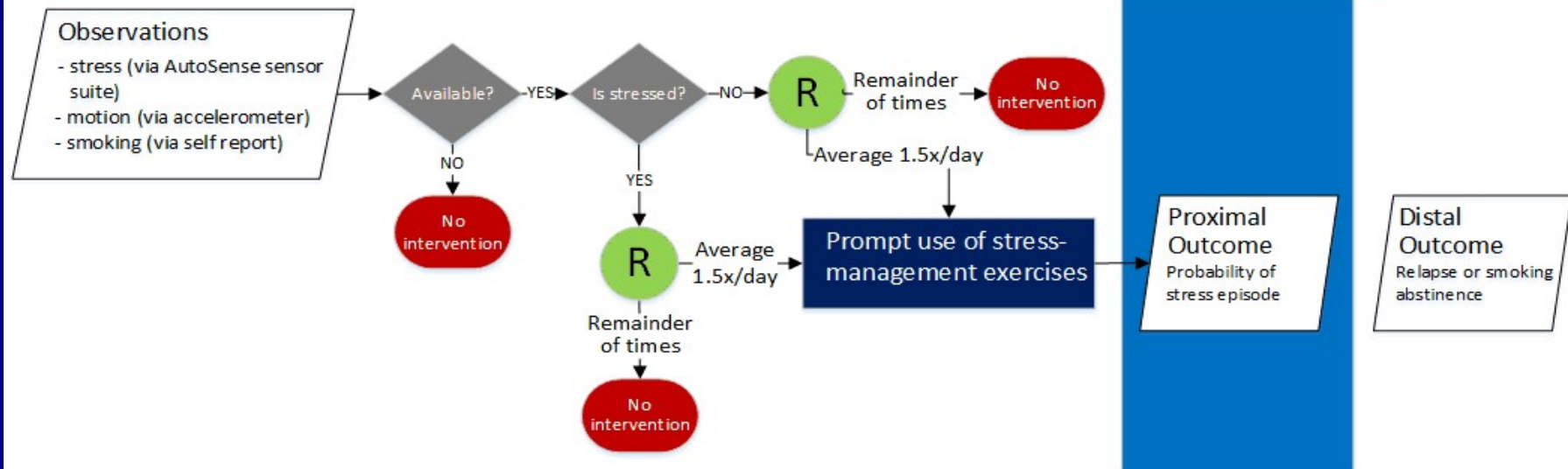
Sense²Stop

Sense²Stop MRT for Stress Management in Newly Abstinent Smokers

Every minute of every day starting with quit date

For two hours
after
intervention is
delivered

Measured via
EMA and
puffMarker over
10 days



PI: S Kumar

Location: Northwestern University, B. Spring, (P.I.)

Funding: NIBIB through funds provided by the trans-NIH Big Data to Knowledge initiative U54EB020404

HeartSteps V1

Heartsteps MRT to Promote Physical Activity Among Sedentary People

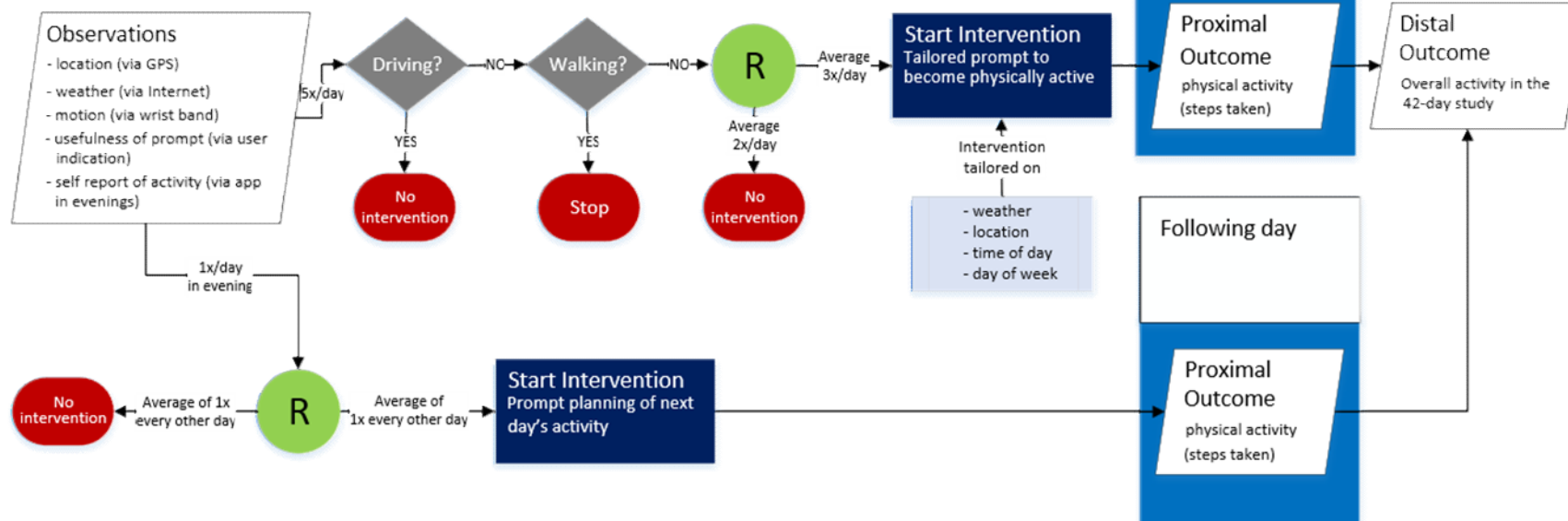
Each day of study

Observations are continuous (except self report)

Randomizations to activity prompts occur 5x/ day at likely times for increasing physical activity

Next 30 minutes
after intervention is
delivered

Measured via
accelerometer
throughout study



SARA

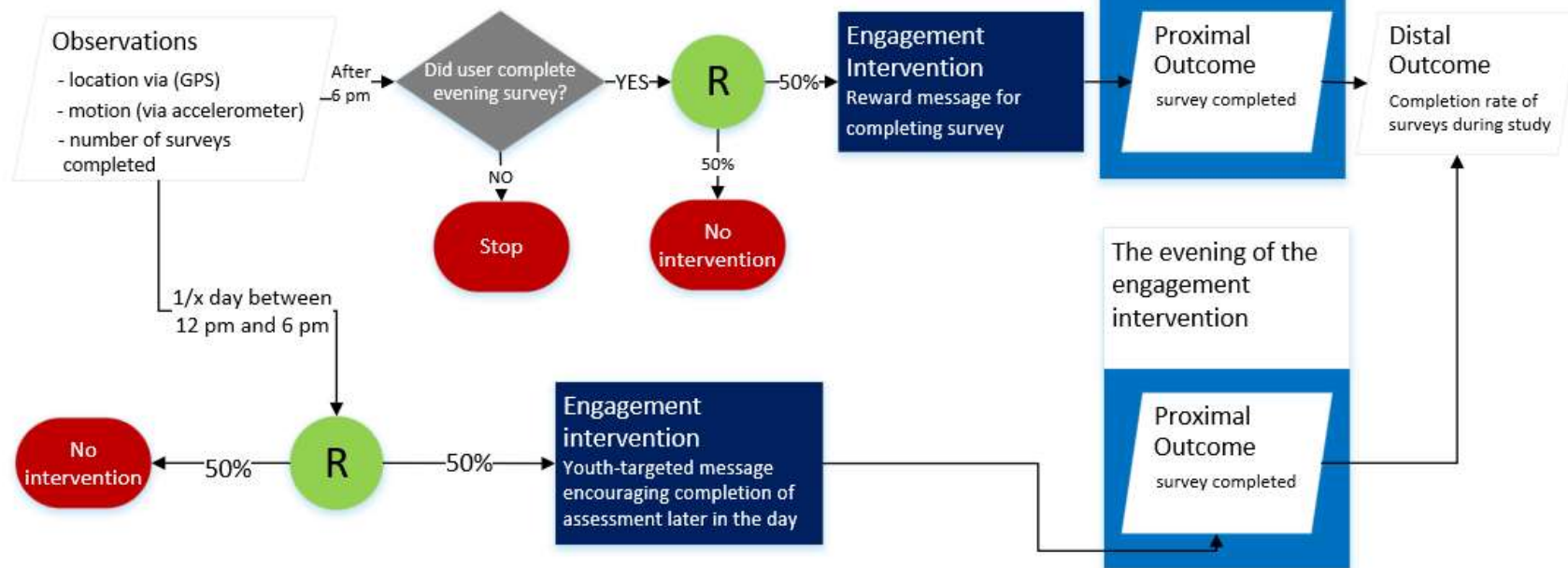
Data Collection MRT to Promote Engagement in Substance Use Research

Each day of study

This chart represents two of four engagement MRTs embedded in the SARA app

The day after the engagement intervention

Day 30



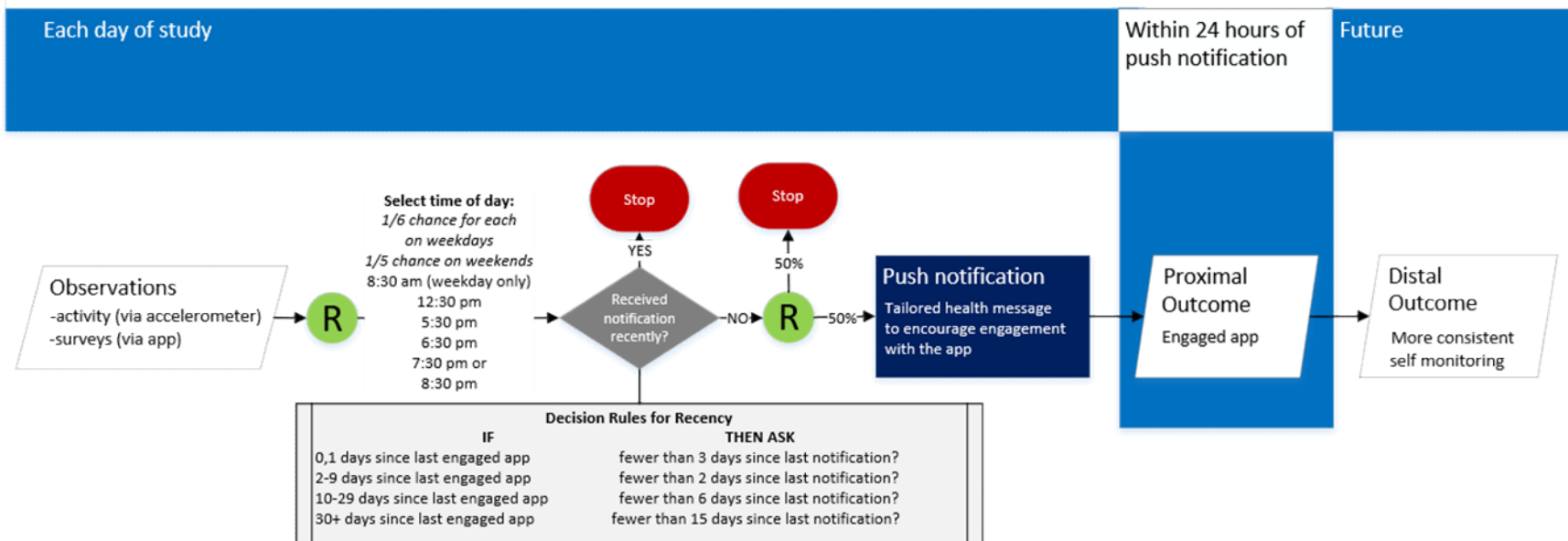
PIs: M Walton, S Murphy, and M Rabbi Shuvo

Location: University of Michigan

Funding: Michigan Institute for Data Science (PI S. Murphy),
University of Michigan Injury Center (PI M. Walton)

Engagement with JOOL

MRT to Promote Engagement with Purpose-driven Well-being App



PI: Victor Strecher, PhD, MPH, CEO of JOOL Health

Location & Funding: Ann Arbor, MI

URL: <https://www.joolhealth.com>