

# Designing clinical trials for therapeutic apps

Upgrading seminar

Lauren Bell  
March 2019

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



# Acknowledgements

I am fortunate to have wonderful supervisors and collaborators for this PhD

I acknowledge the support of the Medical Research Council's Hub for Trial Methodology Research for funding my PhD, including an internship at the BSU-MRC and the support costs to implement a stratified, multi-arm MRT.

Thank you to my supervisors Dr Elizabeth Williamson (LSHTM) and Dr Henry Potts (UCL) for their constant support and dedication to pursuing research excellence

Very grateful to have collaborations with Dr Claire Garnett, Dr Olga Perski, Professor Robert West, Dr Sofia Villar, Professor James Wason and Dr Tianchen Qian.

Institute of Mathematical Sciences at the National University of Singapore, for inviting me to the workshop '*Statistical Methods for Developing Personalized Mobile Health Interventions*'.

1. Disconnect between mobile health body of evidence and uptake of apps
2. '*A Tale of Two Apps*' –case studies of regulatory approval in the US and UK
3. Limitations when sole evidence comes from conventional RCTs
4. The real life implications of these limitations for public health systems
5. Scoping review aims and systematic search terms
6. Examples of digital health evaluated with a SMART, MRT and MOST
7. Findings of the scoping review and possible future research
8. Details of three Work Packages as planned research for the PhD

# Promise of digital therapeutic apps

We are in the middle of a digital revolution!  
By 2020, we expect to have over 6 billion smartphone users  
FDA estimate over half of users would have  
downloaded an app as a health technology.



Promise of digital therapeutic apps includes areas of real-time monitoring, engaging and informing patients with their care and outcomes, treatment adherence, streamlining access to care and communication between carers and patients.

I am primarily interested in digital therapeutic apps for the management of chronic conditions.

# Disconnect between evidence and uptake for health apps

Byambasuren O (2018) found 318,000 health apps are available to the public but only a very small fraction are clinical evaluated .

Rogers MA (2017) discovered that the health apps which are found to be efficacious in a RCT, less than a quarter were found as publically available and functioning.

Has further implications beyond fitness and wellness apps available in app stores.

Digital therapeutic apps are becoming integrated into public health systems to treat depression, diabetes, substance abuse disorder, cancer, schizophrenia and obesity, to name a few.

# Examples of digital therapeutic apps

**Deprexis by GAIA AG** (<https://deprexis.com/>)

**Digital treatment for mild to severe depression. 13 RCT to evaluate effectiveness**

Website says: *Fully automated **treatment adapts to individual patient** needs and designed to fit into the therapy plans of all providers, payers and employers.*

**Reset by Pears Therapeutics** (<https://peartherapeutics.com/>)

**Computerized behavioural therapy device for psychiatric disorders. 1 RCT with contingency management.**

Treatment with medication and contingency management.

Evaluation **did not disentangle the effect** of contingency management and app.

Patients can choose lessons that are relevant to managing their disease or as **recommended by their clinician.**

# Methodological limitations with conventional RCTs

Both Reset and Deprexis are **dynamic** (they change according to the patients outcome over time) and **complex** (they are made up of different treatment components) interventions

We do not know:

**How** effective Reset is, as a component in a complex intervention with (i) contingency management, (ii) medication and (iii) therapy

**How** Deprexis was developed to be personalised – when and how it informs the GP if you not responding well

# Implications of methodological problems

## The Lancet editorial

*Is digital health different?* November 2018

*"The relatively **low barriers to market** entry have encouraged innovative small and medium sized companies, often new to the health market. Research, especially for AI work, remains centred on machine learning outcomes, and the **shift to clinical outcomes** has not kept pace with the products' move into clinical practice. Inherently, **digital products collect a wealth of data in real time**, and **other methods of evaluation might be better suited to this sector.**"*



# Implications of methodological problems

## The Lancet editorial

*Is digital health different?* November 2018

*"Without a **clear framework to differentiate efficacious digital products from commercial opportunism**, companies, clinicians, and policy makers will struggle to provide the required level of evidence to realise the potential of digital medicine. The risks of digital medicine, particularly use of AI in health interventions, are concerning.*

***Continuing to argue for digital exceptionalism and failing to robustly evaluate digital health interventions presents the greatest risk for patients and health systems."***

# Apps learn and evolve

Apps develop differently to pharmaceutical medications

Good practise for app development often begins with a 'Minimum Viable Product', then the app rapidly evolves from knowledge gained with **real-world use and testing**.

Therapeutic apps are **often complex interventions**, with **specific components** that are behaviour change interventions.

There is often no 'one size fits all', thought of as **dynamic, adaptive interventions** which adjust the treatment appropriately as patient's health evolves over time.

# Overarching aim

There is a need for more agile trial designs that synchronise and balance the two separate goals of (i) learning how to continually improve an intervention, while (ii) gathering information to assess overall effectiveness.

Scoping review addresses

***What are the suitable trial designs for the development and evaluation of digital therapeutic apps?***

# Research questions for scoping review

What are the trial designs which are recommended for the development and evaluation of digital therapeutic apps?

Are these trial designs for the development or the evaluation of digital therapeutic apps?

How are these trial designs implemented for the development and evaluation of digital therapeutic apps?

What are the merits and limitations found in theory and practice of each trial design?

And, what are the opportunities for advancing trial methodology for digital therapeutic apps?

**New trials suggested** for digital health interventions, such as the Multiphase Optimisation Framework Strategy, Micro-randomisation and Sequential Multiple Assignment Randomised Trial.

Apps have the ability to **swiftly learn and adapt** from personal real-time data.

Look to:

rapid learning research systems,  
whole phase II/III clinical development programs,  
precision medicine,  
response-adaptive randomised (RAR) trials  
and comparative effectiveness research methods.

Demspey raised a number of considerations and proposals, including:

- (i) that apps can be adaptive and personalised medicines;
- (ii) the important distinctions between micro-randomisation and n-of-1 studies;
- (iii) the challenges with designing trials that are sufficiently powered to detect interactions between components;
- (iv) the potential multi-arm bandits and contextual bandits have in optimising an intervention from real-time data, and;
- (v) testing for the treatment effect conditional on covariate history.

# List of trial designs for systematic review

Sequential Multiple Assignment Randomised Trials for optimising dynamic treatment regimens;

Micro-randomisation for 'just-in-time' push interventions;

N-of-1 and series of N-of-1 for personalisation of apps;

Response-adaptive randomised trials for allocating more patients the most effective app,

Multi-arm Stepped Wedge Cluster Randomised Trials for rolling out components of an app overtime;

Multiple Optimisation Strategy framework and Multi-Armed Bandit Models for building and optimising apps as complex interventions.

## **Inclusion Criteria of studies:**

The digital therapeutic app is an intervention with aims of improving the user's health;

The digital therapeutic app is either standalone intervention or part of a sequence or combination of therapies;

The primary intervention can be delivered on a PC, laptop, tablet or website as well as an app on a smartphone;

The paper can be an original trial, feasibility or pilot study, protocol, statistical analysis plan, simulations of an imagined trial;

Trial outcomes were quantitative or qualitative;

Trial outcomes were efficacy, usability or engagement.



## Exclusion criteria of studies

The digital therapeutic app was only evaluated in a conventional parallel randomised controlled trial without any trials for the development phase;

The app was used primarily for data collection and/or communication of results, rather than an intervention;

The paper introduced the design concept or theoretical methods without an evaluation or simulation.

# Search terms of review – ran November 2018

**This defined selection of search terms to identify interventions includes:** mobile application; mobile health app; mobile health application; mobile app; smartphone application; smartphone app; web-based intervention; mobile health; mHealth; telemedicine; telehealth; eHealth; cell-phone; handheld computer; user-interface and web-portal.

**Our database search coordinated each trial design** (Sequential Multiple Assignment Randomized Trials; Micro-randomisation; N-of-1; Series of N-of-1; Response-Adaptive Randomization; Multiphase Optimization Strategy Framework; Multi-armed Bandit models; Multiarm Stepped Wedge Cluster Randomized Trials) with all the mobile health search terms.

**The searched was conducted in November 2018 with the databases** PubMed, PsycINFO, ISI Web of Knowledge, Science Direct, mHealth and ClinicalTrials.gov.

# Academic database findings

From the academic databases (Pubmed, PsycINFO, ISI Web of Knowledge, Science Direct, mHealth) a total of 45 citations were found. The count of papers by trial design were Sequential Multiple Assignment Randomised Trials (n=4); Micro-randomisation (n=8); N-of-1 (n=12); Series of N-of-1 (n=4); Multiphase Optimisation Strategy Framework (n=11); Multi-armed bandit models (n=1); Response-adaptive randomisation (n=1); Action centred contextual bandits (n=1); Multi-armed Stepped Wedge Trials (n=0).

The results from clincialtrials.gov found additional 14 upcoming trials, not yet published, including; Sequential Multiple Assignment Randomised Trials (n=5); Micro-randomisation (n=2); Multiphase Optimisation Strategy Framework (n=4); N-of-1 (n=2); multi-arm bandit models (n=1).

# Total papers included

In total inclusion was 25 papers, consisting of Sequential Multiple Assignment Randomised Trials (n=3); Micro-randomisation(n=8); Multiphase Optimisation Strategy Framework (n=13) and Series of n-of-1 (n=1).

Common reasons for exclusion:

- (i) Apps which operate algorithms based on bandit models to be adaptive interventions, but only evaluated in a conventional RCT.
- (ii) app as data collection device in the conduct of n-of-1 studies and
- (iii) sequential multiple assignment trials to evaluate a sequence of SMS text messaging

# Sequential Multiple Assignment for Randomised Trials

Trials, with stages over time, to inform the development of dynamic treatment regimes.

Patients are first randomised to initial interventions, outcomes are measured and then based on responses to initial treatment patients are then re-randomised to subsequent treatments.

Key to SMART is the tailoring variable . Success of the SMART design hinges on getting the dichotomisation of tailoring variable right. Ideally need to be valid and good clinical outcome, quickly obtained, not too intrusive or expensive

# Example of a SMART digital health trial

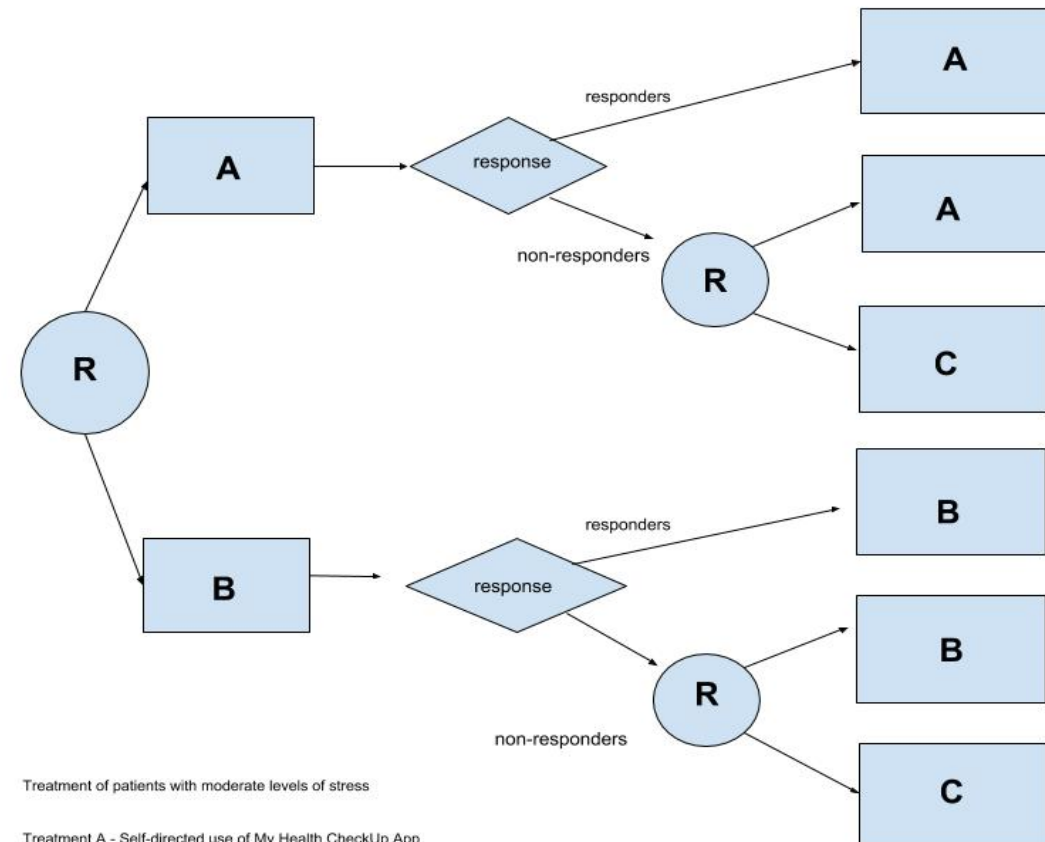
SMART is a study designs which informs the development of adaptive interventions.

## Adaptive Internet-based Stress Management Among Adults With a Cardiovascular Disease: A Pilot Sequential Multiple Assignment Randomized Trial (SMART) Design

ClinicalTrials.gov Identifier: NCT03267953

Intervention is MyHealthCheckUp app

Primary outcomes – feasibility  
Secondary outcomes – clinical



Treatment of patients with moderate levels of stress

Treatment A - Self-directed use of My Health CheckUp App

Treatment B - Minimal Guided use of My Health CheckUp App with layperson telephone coaching

Treatment C - Professional-led motivational interview

Response is measured at 6 weeks

Total treatment duration is 12 weeks

# Micro-randomisation for JITAI

Micro-randomisation is a study design which randomly assigns components within an app, such as tailored notifications, repeatedly to participants during a trial.

A common aim is to optimise time-varying components for complex intervention

Decision times are first established, for example 3pm, 7pm, and 9pm each day. Patients are randomised to receive a notification at fixed probabilities.

Outcomes are Proximal (immediately after the notification) and Distal (after a total treatment period).

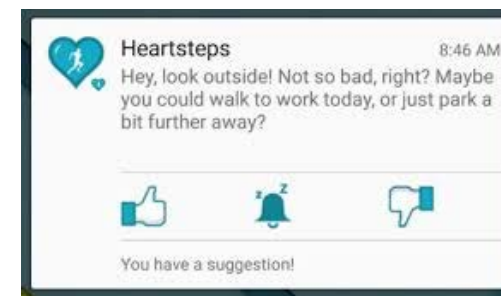
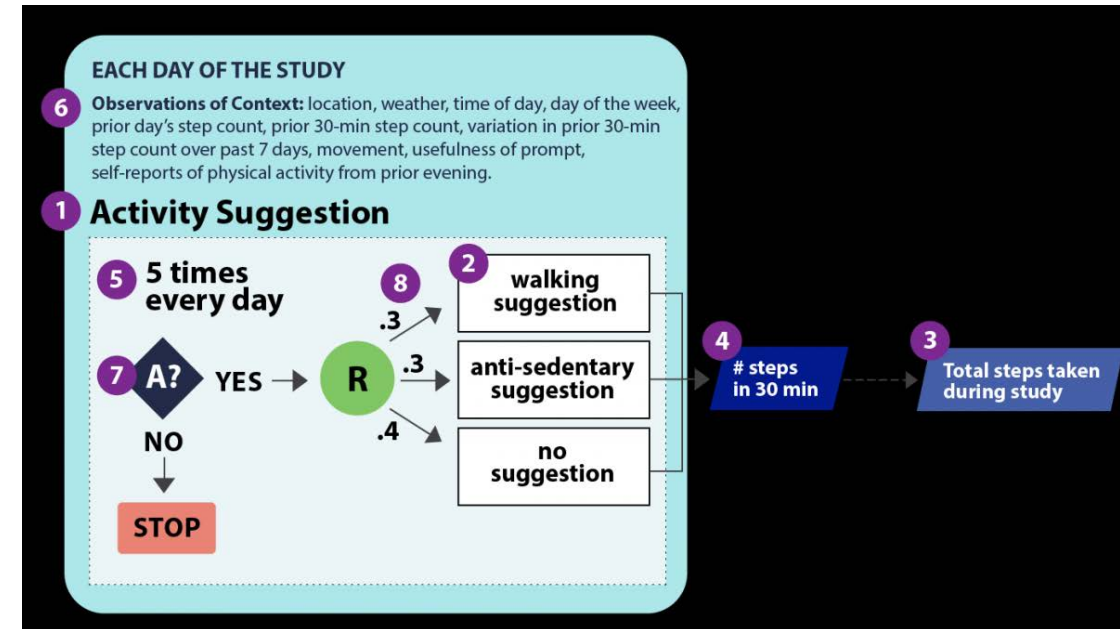
# Example of a Micro-Randomised Trial

*Heart Steps App*, contextual tailored activity suggestions, based on environmental factors.

Randomisation: Five decision time points per day, 0.6 to receive one of two suggestions, 0.4 to receive no suggestion.

Proximal outcome: Total number of steps taken in the 30 minutes following a decision point.

Distal outcome: Total step count during the 42-day study.



<https://methodology.psu.edu/ra/adap-inter/faq>



1. Development outcomes were often efficacy orientated, however, some primary outcomes were also engagement measures.
2. Rabbi leads with a development work to first meet engagement objectives, then follows with adding and testing therapeutic interventions in the next phase. This approach mitigates the risk of testing an app for effectiveness in a trial, only to discover engagement was too poor
3. Methodological research to embed a micro-randomised trial within a randomised found the operational characteristics of type I error rates unlikely to be inflated. However micro-randomisation trials have been development objectives to date – explore the possibility to integrate and synchronize optimising development and evaluation of effectiveness

# Multiple Optimisation Strategy Framework

Multiphase Optimisation Strategy (MOST) is a framework to optimise complex interventions made up of multiple components.

The framework emphasises agile screening experiments of potential components to develop a complex intervention before a definitive RCT.

MOST consists of three phases (i) screening phase, (ii) refining phase and (iii) confirming phase.

The inspiration comes from computer engineering

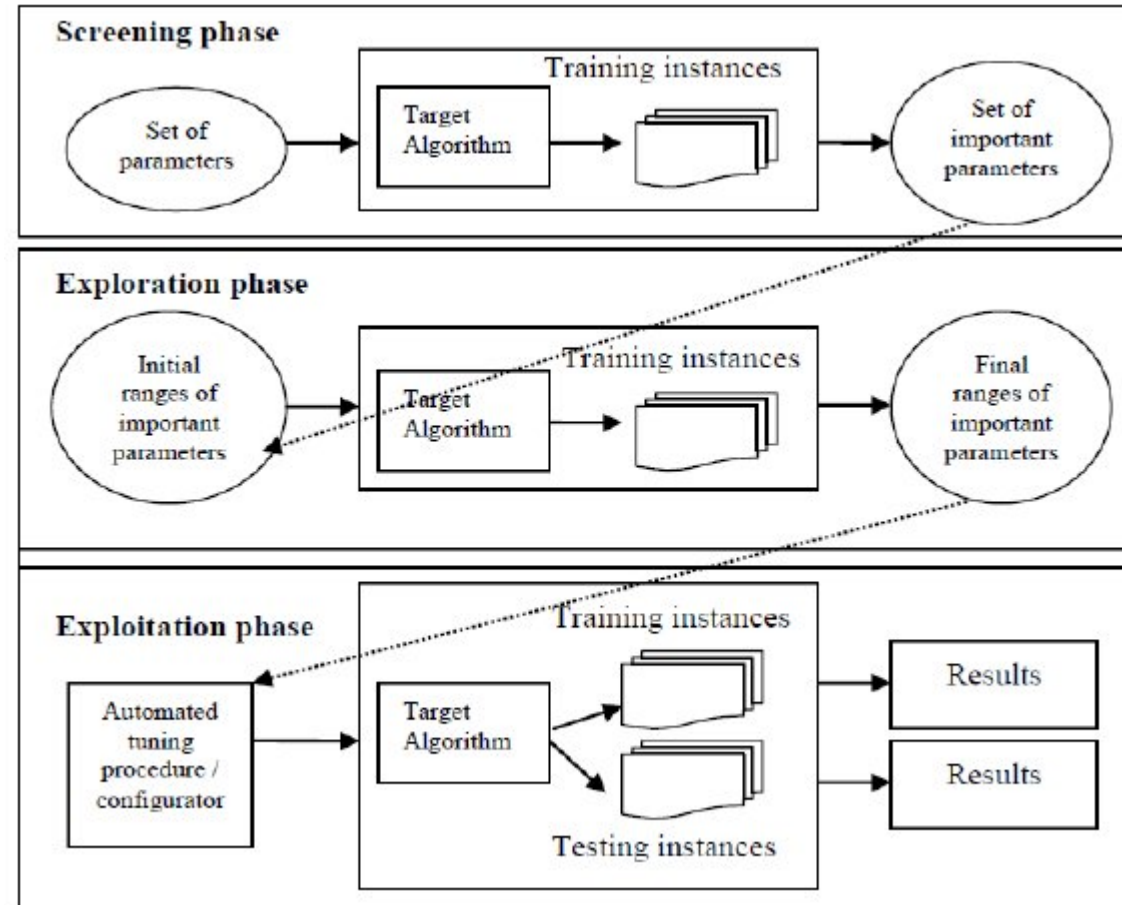
# Inspiration of MOST framework

Conference paper

## ***Fine-Tuning Algorithm Parameters Using the Design of Experiments Approach***

Conference: Learning and Intelligent Optimization - 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011.  
Selected Papers

Aim: to target algorithms and demonstrate that our proposed methodology leads to improvements in terms of the quality of the solutions.

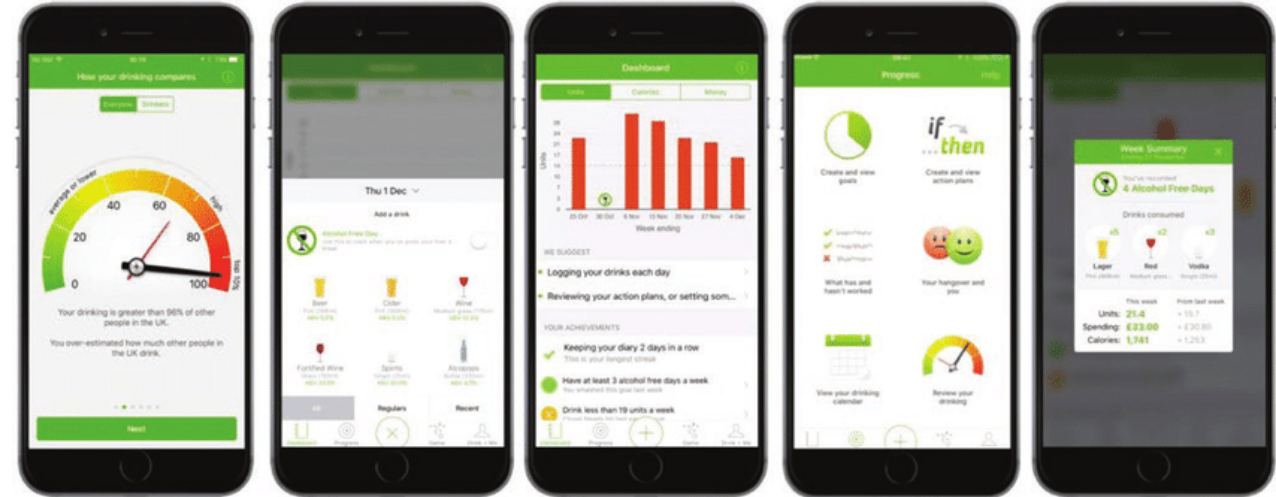


# Example of intervention developed with MOST

## Development of *Drink Less*

Five modules were assessed:

1. self-monitoring and feedback;
2. action planning;
3. normative feedback;
4. cognitive bias re-training;
5. identity change.



[https://www.researchgate.net/publication/324952398\\_The\\_development\\_of\\_Drink\\_Less\\_an\\_alcohol\\_reduction\\_smartphone\\_app\\_for\\_excessive\\_drinkers](https://www.researchgate.net/publication/324952398_The_development_of_Drink_Less_an_alcohol_reduction_smartphone_app_for_excessive_drinkers)

In the refining phase, each module was developed as a 'high' and 'low' version.

This was conducted with a factorial trial with  $2^5$  (32) experimental arms

# Challenges of factorial designs with MOST for *Drink Less*

**No main effect for any component was found, but two interactions were significant, resulting in uncertainty of findings**

*'As no main effect of the intervention modules were found, the significant two-way interactions must be interpreted with caution.'* (Crane 2018).

**Study was powered, but is this the best statistical property for optimising a complex intervention?**

*'This study will recruit 672 participants and have more than 80 % power (with alpha at 5 %, 1:1 allocation and a two-tailed test) to detect a mean change in alcohol consumption of 5 units between the high and low condition for each intervention module.'* (Garnett 2016)

**Follow-up of efficacy data was low.**

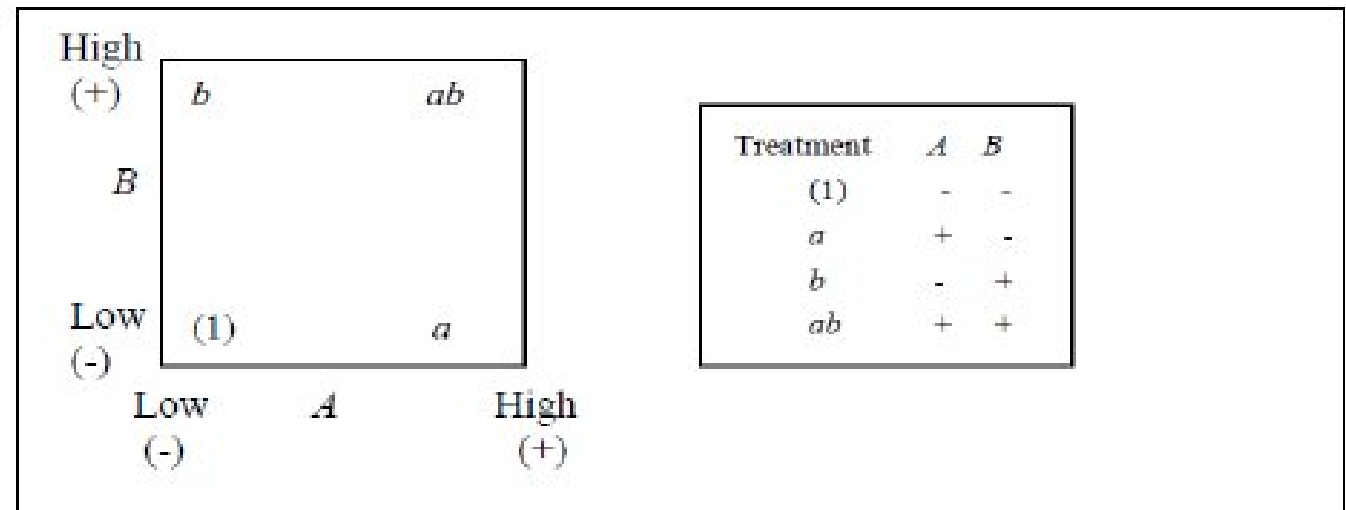
*Of the 672 eligible users, 179 (27%) completed the primary outcome measure at follow-up.* (Crane 2018)

# Factorial Trials

A factorial experiment is a design in which two or more treatments are studied simultaneously.

If the model includes interaction terms, this has implications on

1. Power
2. Analysis

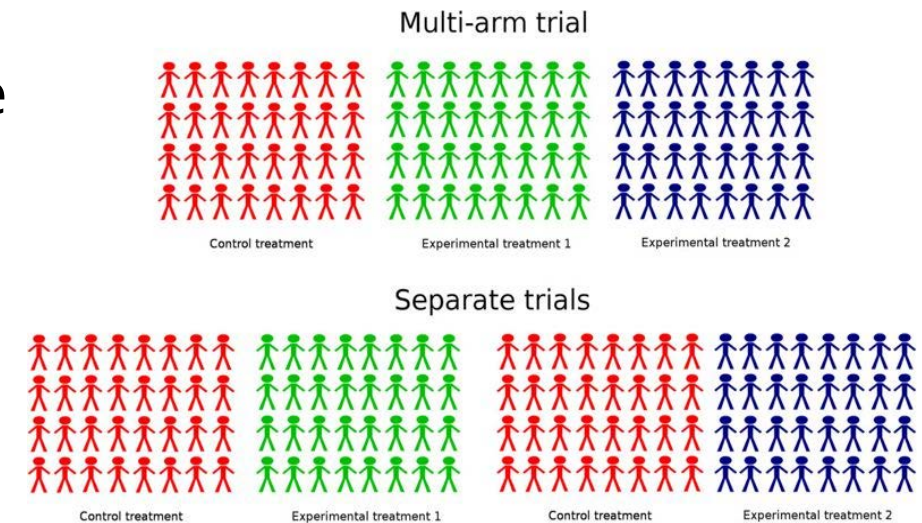


# Multi Arm Trials

Multi-arm trials is a study design in which several treatments, or various combination of treatments, are compared to one control treatment.

Adaptive trial options to increase efficiency include

1. Interim Analysis
2. Response-adaptive Randomisation





# Plans for future research in the PhD

General finding – uptake of the MOST framework and Micro-randomisation trials

Two research questions:

1. The optimisation of mobile health app is a continuous cycle. How can we utilise the available data to help inform the optimal push notification strategy in *Drink Less* through an MRT?
1. Researchers may not be certain there is no clinically significant interaction effects from multi-component interventions – could multi-arm trials, with options of response-adaptive randomisation, stepped wedge clustering and interim analysis provide more definitive results in an efficient manner?



# Future research for the PhD

Analysis of a factorial RCT with *Drink Less* [ $n=179$ ] showed an association between a one-unit increase in total time spent on app (minutes) with an 6% reduction in alcohol use, 95% CI [-0.13 to 0.01].

Previous research by the team suggest engagement with the app is influenced, within the day, by motivation and perceived usefulness.

Work Package One: Understand patterns of use through observational data, define groups with low engagement. We have called these 'Engagement Groups'. Develop interventions and test the optimal delivery and content to increase engagement in these groups

# Objectives of the MRT

## Work Package Two:

*Objectives:* Specific objectives of the trial are:

1. to assess the proximal effect of new tailored push notifications on time spent on the app during the 24 hours following the delivery of the intervention;
2. to explore whether the effect of new tailored push notifications decreases over time, or with cumulative exposure to these notifications;
3. to measure the lagged (delayed) effect of push notifications on the amount of time spent on the app;

Analysis will be done with a centered and weighted least squares estimation method, separately for each engagement group.

## Work Package Three: Alternatives to factorial trials

MOST was designed and analysed with:

(i) Power analysis to detect a treatment effect; (ii)  $2^5$  Factorial design, (iii) Efficacy outcomes with low response to follow-up.

With simulations and theory, we would like to explore if alternative design are more (i) efficient and (ii) provide more definitive results with:

1. Multi-Arm Multi-Stage trials or Multi-Arm Response-Adaptive randomisation;
2. Primary statistical property as the posterior probability the k arm (app) is the most used app (Motivated by Phase II development trials);
3. Use data as an surrogate for efficacy ( which is known in real time and no missing data)

# Considerations for response adaptive randomisation

Randomisation and balance are in conflict, leading to a framework of how we consider the efficiency of a design.

Reflecting on Cox's (1982) comment *'It is little consolation to an experimenter confronted with an unbalanced randomisation to claim that, on average, the randomisation used produces good results.'*

Focus is on the simulations' variability of the allocation proportions, consider the risk a randomisation scheme could performing badly, by examining the distribution of the allocation probability, for different urn based models for different response variables.

Different types of allocation rules: Completely Randomised, Deterministic, Generalised Efron Biased Coin, Atkinson's Rule, Bayesian Rule, Balance Covariates, Balance Covariates with a Biased Coin, Minimization and Minimization with a Biased Coin.

# Analysis and report on the MRT

[illegible]