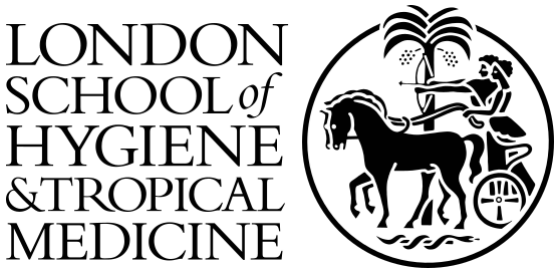


Designing Randomised Trials to Improve Engagement through Optimising the Notification Policy of a Behaviour Change App

Lauren Marie Bell

Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy of the University of London

Funded by the MRC Network of Hubs for Trials Methodology Research
(MR/L004933/2-R18)

Department of Medical Statistics
Faculty of Epidemiology and Population Health
The London School of Hygiene and Tropical Medicine

Abstract

Behaviour change apps can help people maintain healthy lifestyles, however a common challenge for many behaviour change apps is that they lack sufficient engagement strategies to achieve long-term behaviour change. Push notification policies are commonly used within behaviour change apps to maintain engagement over time.

There is a growing recognition of the important factors which impact engagement that go beyond the static features of the app alone. Engagement fluctuates within and between users, and users' varying contextual states, such as their environment, cognitive state or recent individual history, are likely to impact their engagement with the app over time, and in turn the app's effectiveness. Tailoring the notification policy to support individuals, with considerations to such varying contextual states, may help improve effectiveness.

The goal of developing a behaviour change app to be an adaptive, dynamic

intervention presents unique statistical and methodological considerations. Such challenges include understanding the causal effect of time-varying interventions and creating an evidence base for developing decision rules to deliver adaptation. A novel trial design, called the Micro-Randomised Trial (MRT), allows for the estimation of the causal effect of a time-varying intervention, and to inform the development of such decision rules.

The thesis includes (i) a scoping review of randomised trial designs for the development of behaviour change apps; (ii) through data visualisations, an exploration of patterns of engagement with *Drink Less*, a behaviour change app which aims to help reduce alcohol consumption, to inform the design of a Micro-Randomised Trial (MRT); (iii) a simulation study to explore the consistency and efficiency of two estimators to estimate the causal marginal near-term effect of the notification on engagement; and (iv) protocol and findings from an Micro-Randomised Trial (MRT), to both understand and inform the optimisation of the notification policy.

Declaration

Statement of Own Work

I, Lauren Marie Bell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, this has been indicated in the thesis. I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook.

Lauren Marie Bell

26th July 2022

Acknowledgements

Embarking on research to optimise a complex intervention requires optimists.

My deepest gratitude goes to Professor Elizabeth Williamson and Professor Henry Potts for their supervision of this PhD. I am indebted to your commitment, focused attention and dedicated involvement in every step of this work, especially during some challenging times. I am a better researcher because of your guidance, particularly as you taught me how to always strive to explain complex ideas with simplicity and clarity.

I am very grateful to Dr Claire Garnett and Dr Olga Peski for their collaborations with *Drink Less*. Working with you on such a well developed behaviour change app added significant meaning and invaluable experience to the PhD. My sincere thanks also goes to Dr Tianchen Qian for his enthusiasm, kind support and enduring patience, especially as I grappled with the causal inference aspects of the project.

I wish to thank the MRC-NIHR Hub for Trials Methodological Support for

funding this work, along with the administrative staff at LSHTM, Jenny Fleming and Lauren Dalton.

And finally to family and friends. Thank you to my Dad and Juanita, Mum and Richard, to Jo, Colin, Felicity and Oliver, Simon and Cazz, Helen and Kate. To my friends Sarah, Steph, Emma, Krisz, Vicki and Matt, Laura and Jamie. Thank you for listening whenever I needed to talk, and your cheers and encouragement over the years.

And thank you especially to Andrew, May and Avery. This all would never be possible without your devoted love and unwavering belief in me. Somehow you all know the right words to say to me whenever I need to hear it the most.

Acronyms

AIM-ACT Affect-Integration-Motivation and Attention-Context-Translation Framework. 38, 39

AUDIT Alcohol Use Disorder Identification Test. 222, 223

CEE Causal Excursion Effect. 13, 180, 182, 250, 251

COM-B model Capability, Opportunity, Motivation Behaviour model. 37, 38, 43

DAG Directed Acyclic Graph. 41

EMA Ecological Momentary Assessment. 10, 21, 51–53, 62, 76, 113

EMEE Estimator for the Marginal Excursion Effect. 13, 22, 143, 159, 160, 169, 176, 182, 183, 195, 196, 215, 217, 250, 251

EMI Ecological Momentary Intervention. 10, 21, 51–53, 62

FDA Food and Drug Administration. 68–70, 73

GEE Generalised Estimating Equation. 160, 177, 182, 183, 195, 251

IAPT Improving Access to Psychological Therapies. 71, 72

JITAI Just-in-time Adaptive Intervention. 10, 21, 22, 54–58, 62, 64, 115, 116, 255, 256

MOST Multi-phase Optimisation Strategy Framework. 46–48, 77, 78, 82, 106, 108, 269

MRT Micro-Randomised Trial. 2, 13, 17, 21–23, 58, 59, 62, 89–91, 110, 111, 113, 114, 117, 118, 143, 145, 159, 160, 162–164, 166, 167, 169, 171–175, 177, 180, 182, 197, 198, 202, 203, 207, 208, 211, 213, 215, 221, 225, 250–252, 255, 256, 258

MSM Marginal Structural Models. 175, 178, 179

MVP Minimal Viable Product. 33, 43

NHS National Health Service. 29, 64, 71

NICE National Institute for Health and Care Excellence. 71, 205

NIHR National Institute for Health and Care Research. 64, 65

RCT Randomised Controlled Trial. 21, 31–33, 46, 47, 64, 86, 260, 261, 269

SCMM Sequential Conditional Mean Model. 175, 177, 179, 251

SMART Sequential Multiple Assignment Randomised Trial. 77, 266

TDF Theoretical Domains Framework. 37, 38

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

My interest and motivation for this PhD came from my previous post, working as a Statistical Advisor for the Research Design Service (RDS) and Pragmatic Clinical Trials Unit at Queen Mary, University of London (QMUL). My role at RDS involved being the first point of contact for researchers at QMUL or wider NHS services, such as The Royal London Hospital, when seeking statistical advice for grant applications to the National Institute for Health Research, or other funding bodies.

I noticed there was an increasing number of healthcare workers who wanted to evaluate a behaviour change intervention for a chronic condition that was

delivered as an app, however I was not aware of any available guidance on how such interventions should be evaluated. Two current debates became more familiar (i) the suitability of conventional RCTs to evaluate digital therapies such as behaviour change apps, and (ii) the importance of engagement to achieve behaviour change. My RDS supervisor, Professor Richard Hooper, recommended an advertised scholarship with Professor Elizabeth Williamson and Professor Henry Potts, with a position to explore such issues, and I applied.

1.2 Thesis Structure

This thesis is formatted in the style of Chapters that are both published or submitted journal articles (prefixed with "Paper") and more traditional thesis style chapters.

Chapter Two provides a background to behaviour change apps and describes the importance of engagement and personalisation, as discussed in the scientific literature.

Chapter Three describes how dynamic and heterogeneous individual traits can be captured through a behaviour change app, which can also deliver momentary interventions. I introduce Ecological Momentary Assessment (EMA) and Ecological Momentary Intervention (EMI). This chapter also introduces the Micro-Randomised Trial (MRT) and Just-in-time Adaptive

Intervention (JITAI).

The aims and objectives of the thesis are presented in *Chapter Four*, along with an introduction to *Drink Less*.

Chapter Five explores the current challenges regulatory bodies face when evaluating digital therapeutic apps, through two case studies.

A scoping review into the uptake of various trial designs that is provided in *Chapter Six.* These trial designs include MRTs, N-of-1s, SMART and factorial trials within the MOST framework.

Chapter Seven explores patterns of engagement with *Drink Less*. In this paper, I demonstrate how simple visualisations helped gain an understanding of temporal patterns of engagement with *Drink Less*, and how this provided a pathway towards further optimisation of the notification policy to improve engagement.

The protocol for a Micro-Randomised Trial (MRT) is presented in *Chapter Eight*. This trial aims to understand the effect of sending a notification on near-term engagement, and to build an evidence base to further optimise the notification policy for *Drink Less*.

Additionally, *Chapter Nine* includes a description of the Estimator for the Marginal Excursion Effect (EMEE) with simulations to explore performance measures.

Chapter Ten presents the findings of the MRT. This work concludes that notifications are powerful tools to boost 'in-the moment' engagement, however this large near-term effect of the notification does not translate into a policy which keeps users engaged for longer, when compared to a version of *Drink Less* with no-notification policy. I conclude that further optimisation is required, and hypothesise that sending notifications to encourage engagement, only when users are at risk of disengagement, could lengthen the time to disengagement.

Finally, *Chapter Eleven* summarises the key findings from this thesis. I discuss the strengths and limitations of this research, with possible pathways forward to further optimise the notification policy.

1.3 Statement of jointly authored publications

This research is my own work, with contributions from collaborators: Claire Garnett (CG), Yihan Bao (YB), Zhaoxi Cheng (ZC), Tianchen Qian (TQ), Olga Perski (OP), Henry Potts (HP) and Elizabeth Williamson (EW) as follows:

All Chapters were written by Lauren Bell (LB).

Chapter Seven LB was the lead author in the following published paper:

Bell, L., Garnett, C., Qian, T., Perski, O., Williamson, E., & Potts, H. W. (2020). *Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study.* Journal of Medical Internet Research, 22(12), e23369.

LB conceptualised and designed the study with advice from CG, TQ, OP, HP and EW. CG extracted the Drink Less data and gave guidance on how the app was developed and works. LB manipulated and analysed the data with key support from HP and EW. LB prepared the manuscript and wrote the first draft. All authors contributed to and approved the final version of the published paper. LB responded to peer review comments with support from all authors.

Chapter Eight LB was the lead author in the following published paper:

Bell, L., Garnett, C., Qian, T., Perski, O., Potts, H. W., & Williamson,

E. (2020). *Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial.* JMIR Research Protocols, 9(8), e18690.

This publication is the output of a study that LB conducted as part of the PhD.

LB conceptualised and designed the study with contributions from CG, TQ, OP, HP and EW. LB applied for research funds from the MRC Hub for Trials Methodological Research for this study. CG extracted the data. LB analysed the data. TQ designed the sample size simulations. LB prepared the manuscript and wrote the first draft. All authors contributed to and approved the final version of the published paper. LB responded to peer review comments with support from all authors.

Chapter Nine LB was the lead author in the following paper which is accepted for publication in the journal *JMIR mHealth and UHealth.*

Bell L, Garnett C, Bao Y, Cheng Z, Qian T, Perski O, Potts H. W., Williamson E (2022). *Optimising the Notification Policy to Improve Engagement with an Alcohol Reduction App: Results from a Micro-Randomized Trial.* JMIR Mhealth Uhealth. doi: 10.2196/38342.

LB collected and transformed the data and analysed the results. TQ, YB and ZC provided the the code for the estimation of the primary outcome of the MRT. LB wrote the first draft and all co-authors contributed to the final,

submitted manuscript.

Chapter 2

Background

2.1 Digital Health

2.1.1 Types of digital therapeutic interventions

Digital technologies, encompassing wireless systems, sensors and mobile phones, are now ubiquitous across the globe [29]. Over the past decade, smartphones have brought unprecedented changes to our societies, revolutionising economies, cultures, and lifestyles. Smartphones, and software applications that can be executed on them (apps), have changed every aspect of the way we live [43]. This includes how we treat and manage long-term, chronic conditions [65].

The range of digital health apps available to treat chronic health conditions

reflects the recent technological advancements. The term Mobile Health App can refer to any technology which is delivered through a mobile device for public health purposes, and aims to address some aspect of health and wellness [167]. The majority of health apps, which are available freely or for a small fee, aim to promote health behaviours through tracking (such as self monitoring) and self-management [93]. When the app includes various active ingredients that are grounded in behaviour change theory, the app can be known as a behaviour change app. An example of a behaviour change app is *StopApp*, an app which aims to increase the uptake and attendance to NHS Stop Smoking Services. The development of this app included the incorporation of evidence-based Behaviour Change Techniques (BCTs), to target the specific behaviour. One BCT included in *StopApp* is *Comparative imagining of future outcomes*, which encourages users to imagine positive stories of life and positive health benefits they will gain once they have quit smoking [83, 77, 143].

Another class of mobile health apps that is poised to revolutionise health systems is the Digital Therapeutic App [50, 15, 210]. Digital Therapeutic Apps are prescriptive software to treat various conditions [215], embedded into clinical pathways, with such apps available to treat depression [115, 207, 138], diabetes [103, 79, 38], substance abuse disorder [52], cancer [190], schizophrenia [75] and obesity [109]. These apps are often multi-component, complex interventions with specific behavioural change therapy components. Their

functionality builds on the collection of personal data over time, through self-reports, or incorporation with Electronic Health Records. Personal data can also be gleaned through various sensors and WiFi connections, including cameras, microphones, accelerators and cloud storage [93]. Similar to Digital Therapeutic Apps, but perhaps without the behaviour change components, is the Mobile Medical App [131], which is a type of software driven applications that meets the definition of a medical device and requires regulatory approval [166, 74]. These apps are clinical tools used in medical practice, as opposed to a behaviour change app which can be used at home [7].

Management of chronic, noncommunicable diseases

The World Health Organisation estimates that each year, more than 41 million people are killed by noncommunicable diseases, specifically, by cardiovascular diseases (17.9 million), cancers (9.3 million), respiratory diseases (4.1 million) and diabetes (1.5 million) [81]. The risk of such diseases is known to increase due to modifiable, long-term behaviours. For example, the harmful use of alcohol attributes to over 3.3 million deaths a year, tobacco over 7.2 million deaths a year, and insufficient physical activity 1.6 million deaths a year. Public health institutions prioritise health behavior change through the delivery of behaviour change interventions, notably, the longer-term economic viability of the United Kingdom's National Health Service (NHS) depends on successfully engaging its citizens in preventive health behaviour [1]. The delivery of *individual-level interventions* should be optimised for providing

motivation and support to the individuals [1]. An appealing option, with the ubiquity of smartphones, is the development and deployment of behaviour change apps [237].

Behaviour change apps are novel, complex interventions [150]. Such apps are based on behaviour change techniques, and include components that are linked with theories of behaviour change [70]. Understanding the effectiveness of a behaviour change app in reducing harmful behaviours, such as excessive alcohol consumption or tobacco use, is crucial if the app is to have any impact on reducing deaths from noncommunicable diseases [136]. There are broad convergences on the importance of transparency, health content and interoperability, yet there is currently no international consensus on how to evaluate such interventions for effectiveness, regardless of the numerous evaluation frameworks available to do so [123]. This lack of a set evaluation framework makes it difficult for clinicians or patients to identify safe and effective behaviour change apps [72]. Broadly, to understand if a health technology improves health outcomes, evaluation could be done solely through a parallel, randomised controlled trial, such as how medical drugs are typically evaluated, yet the suitability of such an approach for behaviour change apps is debated [174].

The known challenges of evaluating behaviour change apps through conventional, parallel, randomised trials are intricate and considerable. Prevailing issues include (i) the pragmatic, rapid and agile development cycles to de-

liver frequent updates, which is essential for the app to remain relevant and stylish, to ensure usability; (ii) the importance of sufficient engagement to achieve long-term behaviour change; and (iii) the differing, fluctuating needs and wants of individual users, and their changing environment over time. These issues are discussed in detail in the following sections.

2.1.2 Experimental designs to match the rapid, agile development cycles and frequent updates of behaviour change apps

Generally, behaviour change apps have immense appeal with patients, due to remarkable accessibility, personalisation and scalability, yet for all this potential, we face a significant disconnect between the behaviour apps marketed and accessed, and the extensive body of research into evaluating behaviour change apps. Two reviews show this disconnect stems for multiple sources, as more than 318,000 mHealth apps are available to the public but only a very small fraction are clinically developed and evaluated [37]. Furthermore, of the mHealth apps that are shown to be efficacious through a Randomised Controlled Trial (RCT), less than a quarter were found to be publicly available and functioning [192].

The Randomised Controlled Trial (RCT) is the established gold standard of evaluation methods to determine if an intervention causes the improvement of health outcomes [96]. Confounding is commonly present in observational

studies, with a confounder being a variable associated with an exposure and an outcome, which creates a non-causal component association between the exposure and the outcome. Randomisation eliminates any confounding by breaking the links between exposures and any potential counfounders. That is, randomisation gives strength to research designs by creating comparable or exchangeable groups of individuals, randomly allocating a group of individuals to receive an intervention, and a different group of individuals to receive a comparative treatment, such as a placebo or control. Randomisation is a study design property to best reduce bias and confounding for causal inferences [199].

However, RCTs present methodological challenges when evaluating behaviour change apps [158]. To ensure internal validity, RCT protocols generally require the intervention remain the same, or 'frozen' during the entire trial, denying the developers the ability to continuously improve the app. This requirement, coupled with time frames of RCTs (RCTs take, on average, from enrolment to publication, 5.5 years to complete [106]), generally means that any technological advancements during the trial cannot be integrated into the app. Freezing the intervention, for internal validity purposes, could render the app too dated and 'clunky', impairing usability, resulting in the intervention becoming obsolete.

There is a need for study design frameworks to fit the agile, iterative cycles of development and evaluation commonly implemented in the digital

industry. A framework for digital development, known as a lean start-up technique, is the Minimal Viable Product (MVP) [188]. This framework reflects a *learn-as-we-go* process, testing incremental changes with cycles of research, development, and evaluation [130]. The MVP framework begins with rolling out a product with minimum features, and then allows developers to test if further features should be included, separately over time, to quickly understand if the app is engaged with as intended.

Design features of the app, to incrementally improve engagement through improving the ease of interface, visual appeal, warmth, and appropriate friendly support can be tested, through patterns of behavioural engagement in MVP cycles. In the digital industry, MVP mitigates the risk of exhausting resources set aside to develop a behaviour change app, prior to any evaluation, to then find the app was not engaged with as intended and, in turn, found to be an ineffective intervention in an RCT.

2.1.3 The differing, fluctuating needs and wants of individual users and their environments over time

A common finding from adherence and behaviour change theory research is that there is often no *one size fits all* [69], that is, subsets of clinical populations have different needs and wants from a behaviour change app, and these different wants and needs change over time. The personalisation of a behaviour change app means the app adapts to an individual's changing

states, wants and needs over time [24]. Developing such 'intelligent' behaviour change apps, that are personalised interventions, could improve the perceived usefulness, and in turn engagement over time, with the behaviour change app.

2.2 Engagement

Poor engagement is considered a key reason behaviour change apps may fail to be effective interventions in conventional RCTs [79, 80]. Engagement, a construct of both behavioural and experiential aspects [171], gives insight to the experience and comfort users have with the intervention.

Engagement is studied in various fields, and commonly plays a fundamental role in interventions within public health, digital health, human-computer interaction, marketing and education [160]. Engagement can be thought of as a state, which exists within the dynamic, real-world settings, with digital interventions competing for an individual's attention alongside the multiple needs and external stimuli that change over time [160].

Arguably, there typically needs to be some level of engagement with an intervention for any attributable effectiveness, however the association between engagement and effectiveness is unlikely to be a simple linear dose-response relationship. That is, more time spent on the app may not directly translate into an increased improvement in outcomes, and efforts to always maintain

or increase time spent on the app may overburden or annoy the user.

2.2.1 Definitions of Engagement

Perksi [171] defined engagement to be made up of two broad aspects, behavioural and experiential. Under this conceptual framework, engagement with an app can be partitioned into the temporal patterns of usage, such as the frequency, amount, depth, and duration of use, and the experiential engagement, thought of as a state of 'flow', a cogitative state involving enjoyment and focused attention with the intervention [171].

A recent review by Nahum-Shani [160] further expanded the definitions of engagement to be an "energy investment directed by an individual toward a focal stimulus or task" and effective engagement to be "the extent, frequency and duration of investment of physical, cognitive, and affective energies in focal stimulus or task needed to bring about a pre-specified outcome". That is, the patterns and extent of engagement can be thought to be effective only if the ultimate, distal goal is achieved. This distal goal may be an improvement in the patient's health. Effective engagement is defined in relation to the pre-specified, overall outcome. Nahum-Shani's review also expands engagement to be either positive and negative, with engagement as "*state of energy investment involving positively (vs. negatively) valenced physical, affective, and cognitive energies directed toward a focal stimulus or task*". Examples of negative engagement can include clickbait. Clickbait is a strategy to attract a

user's attention but ultimately is an unfulfilling or misleading experience, which reduces the perceived usefulness of the application. Another concerning negative engagement experience would be trolling or bullying on social media. Behavioural engagement with an intervention may not always result in a positive, experiential experience.

Another aspect of negative engagement is the exposure to repeated stimulation, which can cause habituation [186]. When a person is subject to habituation, the magnitude of their response to a specific stimulus reduces over time [91]. An example of habituation observed in a research study was office workers who began to ignore repeated security warnings over time [220]. To counteract such risks of habituation, recovery periods can be implemented, when a user is given a period of rest from the stimulus [220].

Individual-level factors which influence engagement can be social, psychological and neural. Engagement, a multifaceted construct of various forms, is a broader, more complex term than adherence with an intervention. Adherence is the act of undertaking a requested task or following prescribed treatment regimes [151], and can be affiliated with behavioural engagement (i.e., use of the app) [171]. Adherence in a behaviour change app can be self-monitoring when prompted, where behavioural engagement (i.e., use of the app) may be operationalised as a user recording their recent alcohol units consumed (regardless of being asked to or not).

The holistic definition of engagement, beyond adherence, includes the as-

pects of cognitive and attention of the task, involving the self-reflection of their drinks consumed to align with their overall goals to cut back on drinking [160]. Behavioural engagement provides some indication of how the app is engaged with, but does not comprehensively represent the concept of engagement. This is because behavioural engagement (i.e, use of the app) as a metric omits the experiential engagement components.

2.2.2 Theoretical frameworks

When developing or optimising a complex intervention, it is important to map the hypothesised mechanisms of actions to show how the intervention brings about its effects. This can be formulated in a framework of a theoretical model. The Theoretical Domains Framework (TDF) [41] is a integrative framework of behaviour change theory, to help make the theory more accessible and usable for other disciplines. The framework comprises of 14 domains of theoretical constructs: Knowledge, Skills, Social/Professional Role and Identity, Beliefs about Capabilities, Optimism, Beliefs about Consequences, Reinforcement, Intentions, Goals, Memory, Attention and Decision Processes, Environmental Context and Resources, Social Influences, Emotions and Behavioural Regulation. An example of such a theory to understand behaviours is the Capability, Opportunity, Motivation Behaviour model (COM-B model))[149], which maps key contextual factors to the Theoretical Domains Framework (TDF). The COM-B model relates the psychological states of capability (i.e., skills and knowledge), motivation (i.e., reflec-

tive and/or automatic) and opportunity (i.e., social influences) as pathways (i.e., moderators) for behaviour change.

An example of the development of a behaviour change app, based on the COM-B model and TDF, is an app which aims to change the behaviour of parents for the weight management of their children (developed by Curtis et al in 2015 [60]). The development of the childhood weight management app, for parents to use, utilised the COM-B model and TDF to explore the barriers and facilitators parents experience, in terms of their capability, opportunity and motivation to provide their children with healthy food portions. Key secondary goals towards the development of this app were focused on social validity (i.e., acceptability amongst its stakeholders) and engagement [60]. The results of this research included features of the app which teach parents how to measure portion sizes with graphical tools when preparing meals, and this feature was informed by the TDF of *Skills (cognitive)* and the COM-B model components of *Psychological capability*. Another example within the app, is the feature which utilises every-day household objects (i.e., cups or plates) to correctly measure healthy portion sizes. This component restructures the home environment to increases the capability and opportunity of parents to create healthy portions at mealtimes. This feature was informed by the COM-B model of *Physical opportunity* and TDF of *Environment context and resources*.

Recent developments also includes the Affect-Integration-Motivation and Attention-

Context-Translation Framework (AIM-ACT) framework [160]. This framework considers how the personal neurophysiological states can influence positive and in-the-moment engagement. Three elements are incorporated into the framework: attention, context and translation (ACT) of motivation to behaviour to help guide stable and dynamic engagement features to increase the effectiveness of the intervention.

2.2.3 Engagement as a Mediator and a Moderator

Cognitive behavioural theories can guide the development of the dynamic aspects of when and how engagement is encouraged to occur. For example, theories can guide how various momentary environments may influence the physical opportunity to engage with an app, or how the personal capabilities of information processing, memory and attention influence the user's psychological capability for any engagement and in turn behaviour change. The fluctuating changes of a user, such as their recent behaviour, their varying clinical states or feeling states and their external environments are all key considerations in the development phase to ensure engagement positively boosts behaviour change [33]. These considerations are commonly dynamic, individual traits.

How engagement relates to the effectiveness of a behaviour change app can be distinguished in the terms of mediation or moderation. As defined by Baron and Kenny [11], a **moderator** can be conceptualised as "the function of a

third variable which partitions a focal independent variable into subgroups that establish its domains of maximal effectiveness in regard to a given dependent variable", and a **mediator** is conceptualised as "the function of a third variable, which represents the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest".

I now relate the above concepts of a mediator or moderator to engagement in which the distal health outcome is achieved through a behaviour change app. Engagement is a complex construct, which encapsulates various forms, including experiential (i.e., feeling states) or behavioural engagement (i.e., use of the app). Building on this definition of engagement, I have developed a conceptual model which places behavioural engagement as a mediator and experiential engagement as a moderator. I argue that the effect of the digital health intervention on a health outcome is *mediated* through behavioural engagement. Some behavioural use of the app is required if the app is to have any effectiveness. That is, the mechanisms of action by which the behaviour change app causes an improvement in health is behavioural engagement with the intervention's theoretically-informed, evidence-based modules. I see behavioural engagement as a factor that is a generative mechanism which influences the behaviour change; i.e., behavioural engagement sits on the causal pathway between the app and an improvement in health outcomes. I then extend this model, and place experiential engagement as an effect modera-

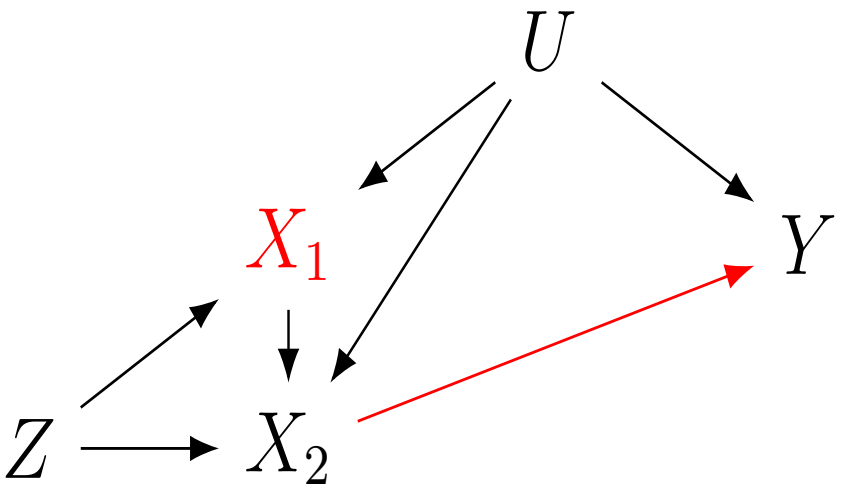
tor on both the behavioural engagement and health outcome. The cognitive and neural states during engagement is an effect-measure modification on behavioural engagement, which can magnify or dilute the strength of the mechanisms of action which take place during behavioural engagement and leads to an improvement in health outcome.

An example of such a theory is how a user's current environment, recent health history, or mood may negatively or positively impact the cognitive and neural states of an individual, which influences the experimental engagement and this in turn modifies the strength of the behavioral engagement to impact behaviour change. For example, a user of a behaviour change app, in an work-place environment, may face competing demands for their time and heightened stress levels. With such competing demands in a workplace, this environment could reduce the cognitive resources to focusing on the app for this moment, and this in turn reduces the behavioural engagement (i.e., length of time and breadth of time spent on the app), which reduces the effectiveness. A different imagined situation is that the person is in a more relaxed environment, with no immediate demands on their attention, which gives more opportunity to be immersed with an intervention, and this increases the length of time and breadth of engagement with the app, cognitive investment and interest, which increases effectiveness.

The following Directed Acyclic Graph (DAG) conceptualises how engagement affects the behaviour change in the above example. The red colour of X_1

and the red arrow represents effect moderation of X_1 on the causal pathway between X_2 to Y.

- Z_1 is *Behaviour Change Intervention*
- X_1 is *Experiential Engagement* (Cognitive and affective energies)
- X_2 is *Behavioural Engagement* (Use of the app)
- U are *Individual states and perceptions* of the app
- Y is *Outcome of an improvement in health*



2.3 Summary

I have established that digital therapeutic apps have great promise and potential to be effective behaviour change interventions. This chapter began by

introducing examples of the different classes of digital health inventions, including behaviour change apps, digital therapeutic apps and mobile medical apps. The range of different classes of digital health interventions is rapidly expanding, as the information and history available to glean through sensors and wearables grows over time, systems link with one another and their potential as paid health technologies in clinical care pathways develops.

The key challenges to deliver on such potential include (i) ensuring their rapid development can be accommodated in any evaluation framework and (ii) the development is based on theoretical frameworks.

The development of digital health interventions can take place with iterative cycles of research, development and implementation, similar to MVP cycles. I described how social, environmental, psychological and neurological aspects are thought to influence 'in-the-moment' engagement, and that a common finding is that people have different needs and wants from their behaviour change app, and these needs and wants can change over time. I described risk to longer-term engagement with repeated stimuli, such as habituation.

Some background on theoretical frameworks is provided, including the Capability, Opportunity, Motivation Behaviour model (COM-B model) and an example of how this theoretical model informed the content development of an app to help support parents in understanding portion sizes and creating healthy meals for this children.

I then detailed two established definitions of engagement, describing how engagement as a multifaceted construct can be a dynamic state. Engagement is not only influenced by the design and features of the intervention but also a user's environment, cognitive capabilities and feeling states. The facets of engagement, which make up the complex, holistic construct, include behavioural and experiential, with further definitions for positive and negative engagement, and for effective engagement. Effective engagement is defined explicitly in terms of the engagement required to achieve the distal outcome (i.e. a improvement in a user's health).

To conclude, I hypothesised that behavioural engagement can be thought of as a mediator to an improvement in health outcomes, and I separated out the experiential facets of engagement as effect-measure moderators, which either magnify or dilute the direct effect that use of the app (behavioral engagement) has on an improvement in health (i.e., the distal outcome).

I will now discuss how complex interventions, such as behaviour change apps, can be developed through a strategic framework which focuses on optimisation of the complex intervention before evaluation. I will then consider how the dynamic, heterogeneous traits can be integrated into the functionality of behaviour change apps, and corresponding research designs, through momentary intervention designs and assessment approaches, to both improve their effectiveness and facilitate evaluation.

Chapter 3

Optimisation Frameworks, Intervention Designs and Assessment Approaches

In the previous chapter, I demonstrated how behaviour change apps are usually some kind of complex intervention, with multiple components and active ingredients. I will now describe and detail some frameworks and research methodologies to develop complex or dynamic behaviour change apps.

3.1 MOST - a framework for optimising complex interventions

The Multi-phase Optimisation Strategy Framework (MOST) is a framework, inspired by an approach in the engineering field, for developing, optimising and evaluating multi-component behavioural interventions [53]. Without a framework, the development of complex interventions may be somewhat arbitrary, with researchers selecting intervention components based on little to no theoretical grounds or empirical evidence.

MOST is neither a particular study design or a procedure, but a framework which emphasises optimisation of a complex intervention before testing the intervention in a resource-intensive RCT. Conventional RCTs can be highly resource intensive due to the time, cost and expertise required to run well [174], with many RCTs failing to recruit enough patients or be completed within budget. The MOST framework allows the intervention to be optimised with the practicalities and insight into real world implementation, to test complex interventions that are optimally developed to be effective, affordable and scalable, before testing in a evaluative RCT [94]. The objectives of MOST aims to examine not only effectiveness, but also address questions regarding the affordability, scalability and efficiency of the complex intervention. Testing an optimal intervention in an RCT not only champions the science, but also the practical aspects of running a trial such as recruitment

and duration of the trial, as more patients should be more willing to be in a RCT which tests a more optimised intervention.

The MOST framework advocates for three phases to optimise an intervention: the *preparation* phase, the *optimisation* phase and the *evaluation* phase. During the preparation phase, researchers typically review the existing literature in the relevant field, and develop a detailed conceptual model that represents the mechanisms of action which the complex intervention would create. The main extension that the MOST framework made to the development of a complex intervention was to add a *optimisation* phase before the *evaluation* phase.

The optimisation phase allows researchers to explicitly model the effects of the different components, leveraging the advantages of effects estimated from randomised trials (to facilitate unbiased estimates for comparative purposes) and the suitability of intervention components for the targeted population. The MOST framework begins by specifying the optimisation goal in the preparation phase. For example, researchers may state their objective is to develop a complex intervention to help participants cut back on smoking, such that (i) the overall cost of providing the intervention is no more than £500 per participant, and (ii) participant burden is no more than 1 hour of their time once a week.

The preparation phase will detail the conceptual model of how the mechanisms of action within the intervention will reach the stated objective [53].

The preparation phase is developed through the theoretical domain frameworks, as described in the preceding chapter. The optimisation phase aims to quantify the effects of individual components and any interactions between components, for the proposed intervention assembled in the preparation phase [53].

The optimisation phase is commonly conducted through a factorial trial or another type of sequential trial to separate out the specific effects [53]. In this context of MOST, factorial trials are commonly undertaken in the optimisation phase for a complex intervention to determine the marginal effects of different therapeutic components.

An example of a digital intervention developed through the MOST framework is the Opt-IN study [206]. The study's primary aim was to select behaviour change intervention components for a treatment for obesity, to maximise individual weight loss, and the secondary aim was to build a complex intervention which would cost under $500. The package is not an app, but is remotely delivered and made up of technology support components.

The preparation phase identified the mechanisms of action to achieve weight loss as self-efficacy, self-regulation, supportive accountability and facilitation. Five intervention components were developed in the preparation phase: (1) Coaching Calls (either 12 biweekly or 24 weekly), (2) Progress Report to Primary Care Physician, (3) Test Messages, (4) Meal Replacement Recommendations and (5) Buddy Training [206]. Then, in the optimisation phase,

a factorial experiment randomised 562 participants to one of 32 experimental conditions was conducted. Measurements for outcomes were taken at baseline, 3 months and 6 months of treatment. The analysis strategy included effect coding. Effect coding is when the variables are assigned a value of either -1 (for the *off* component) or +1 (for the *on* component). The analysis model was a mixed linear model with an unstructured variance-covariance matrix for repeated measures. The analysis strategy first explored 2-way time by component interactions. If the p-value for such interactions was under 0.10, then the component was tentatively selected for the screening stage. The only interaction effect estimate meeting this criteria was the buddy component by time. After the results were considered with the cost specifications, the final components up for consideration, from the optimisation stage were buddy training, 12 coaching calls and PCP reports [206].

3.1.1 Effect coding or dummy coding

It is important to ensure the analysis of data is appropriate for the research questions it aims to address. When analysing effects from a factorial trial for optimisation purposes, the researchers are often interested in estimating main effects of the single component, averaged across all intervention components, as well as any interactions or synergies between two or more components.

There are two common types of coding used for such analyses [53]. *Dummy coding*: the variables are coded as +1 for the *on* component and 0 for the

off component and, *Effect coding*: the variables are coded as +1 for the *on* component and -1 for the *off* component. In standard regression models with dummy coding, in which one or more interaction terms are present, the main effect coefficient for one variable can be interpreted as the effect of increasing that variable by one unit when the other variables are set to zero. Therefore, dummy coding naturally estimates effects of components in the absence of other components. Effect coding allows for the main regression coefficient for a component to be interpreted as the average treatment effect, averaging over the distribution of other components.

3.1.2 Capturing and adapting to dynamic individual states

Complex interventions are sometimes static interventions, and are often optimise from an assemble of static components. That is, the complex intervention remains fixed over time, and does not learn about the individual and adjust the treatment according to their current or recent states.

However, we know that people with chronic illnesses or cyclical conditions, such as diabetes [35], obesity [214, 71], or smoking [47], often experience their conditions or behaviours to change over time. Often people go through different stages of managing their conditions, especially as individual states and environments fluctuate over time. Treating chronic conditions with a more dynamic intervention, which can adapt to such individual states or

environmental changes over time, may lead to developing a more effective intervention [44].

The desire to study the dynamic, heterogeneous traits within populations and their natural environments, have inspired the incorporation of Ecological Momentary Assessment (EMA) and Ecological Momentary Intervention (EMI) into mobile health technologies [99]. In the following section, I describe how EMA and EMI can be incorporated into behaviour change apps to capture the momentary changes within individuals over time.

3.2 Ecological Momentary Assessment (EMA)

As previously discussed, behaviour change apps can capture the dynamic aspects of individuals, such as how the psychological, behavioural and physiological states vary in a person's natural environment. A research approach to formally capture such changes is the Ecological Momentary Assessment (EMA) [201]. The EMA is a research methodology, which uses repeated sampling to capture the temporal changes within individuals over time.

Ecological Momentary Assessment (EMA) is a repeated assessment method to capture behavioural outcomes of individuals as they go about their everyday lives [208]. A few decades ago, EMA studies would often be carried out through paper diaries with pagers, but are commonly implemented through apps now [51]. Such studies allow researchers to study daily processes which

influence behaviour in real-time [201]. States which can be measured over time include internal factors, such as recent drinking patterns and current moods, and environmental cues, such as day of the week or a change in location. The aims of an EMA may be to examine the associations between a person's natural environment and their momentary behaviour or varying symptoms of a disease.

EMA studies are implemented through sampling over time. Research questions may seek to understand behavioral changes over time or behaviour modifications due to different contextual circumstances. As such, the timing of the EMA sampling requires careful consideration, to reduce recall bias and missing data, to maximise ecological validity of the results and minimise habituation to stimuli of the repeated assessment.

3.3 Ecological Momentary Intervention (EMI)

Parallel to EMA is Ecological Momentary Intervention (EMI). Heron and Smyth define EMI as "treatments that are provided to people during their everyday lives and natural settings" [99], with a literature review finding that EMI are scheduled to be delivered at either fixed, random or tailored times. A recent review found that most EMI are now delivered on smartphone apps [9]. EMIs can be synchronised with data gathered over time through an EMA, and can be characterised as "extending the methodology of repeated within-environment prompting into the domain of clinical intervention" [17],

with the idea to provide accessible therapeutic support to individuals when they need this the most.

A review by Ellen Beckjord and Saul Shiffman explored the use of EMA and EMI specifically for interventions to minimise harmful alcohol use [14]. A known barrier with face-to-face interventions, to help people with substance use disorders, is that the skills discussed to maintain safe levels of alcohol consumption during face-to-face settings require the patient to call upon such skills in the moment, when they are most vulnerable to hazardous or harmful drinking episodes [14, 200]. Such skills include problem solving techniques to avoid high-risk situations or interventions to motivate attempts to reduce drinking levels after a relapse. The synchronisation of EMA with EMI offers the opportunity to provide momentary support when the patient is faced with such decisions [142]. Examples of EMA questions include "Are you drinking alcohol?", or "How strong is your urge to drink?". Beckjord and Shiffman argue EMA and EMI should be used together to develop interventions to reduce alcohol consumption, with key challenges of accurately reporting both outcomes and any contextual details.

The contextual factors can be conceptually associated with events of alcohol consumption in a sequential manner, such as (i) the contextual factors which preceded drinking, such as stress or a negative mood, (ii) the environmental factors which influenced drinking alcohol at that moment, such as drinking with friends at a pub or (iii) the consequences after a harmful drinking episode

(such as hangovers and low mood). As dynamic behaviour change theories develop, interventions can be tailored to the individual's dynamic history or past contextual variables to match the mechanisms of action.

3.4 Just-in-time Adaptive Intervention (JITAI)

Just-in-time Adaptive Intervention (JITAI) can be considered as a type of extension to EMI [223]. They are interventions which aim to provide the right support to the right person at the right moment, by adapting the sequence and content of interventions to an individual's changing internal and contextual state [162]. Contextual information about an individual can be gathered through wearable devices, self-report data or monitors. JITAI leverage the collected personal markers as well as predictions of risk to personalize the choice of treatment to the individual.

3.4.1 The concept of timing in a Just-in-time Adaptive Intervention (JITAI)

The components of a JITAI are adaptive, with the aim to intervene when it is most beneficial for an individual given their contextual circumstances. Timing has a different interpretation to that of "clock time", such as 11am or 8pm [162]. Timing in a JITAI is conceptualised as the *state* which an

individual is in at the moment, with the aim to deliver an intervention neither too late nor too early [104].

JITAI requires constant monitoring to understand how an individuals' state evolves over time. Delivering a JITAI is done through decision rules, and means an algorithm decides if the individual is in a state that may benefit from an intervention [159], the type of intervention which might be suitable for the individual in that moment, and if the cumulative sequence of repeatedly intervening has the potential to disrupt or impact the process by creating a negative impact on future treatments and states. That is, JITAI aim to not only influence a short term outcome, in that moment, but a long term outcome of what the overall goal is. This long term goal is known as the distal outcome.

The concepts and components which make up a Just-in-time Adaptive Intervention (JITAI) include [162]:

3.4.2 Concepts

State Dynamic, multi-variables that represent the 'space' of an individual [97] when a mechanism may produce an effect. The dynamic state variables may be made up of past state values of an individual, external events or internal processes.[10]. An example of a state could be if the individual's feeling state of the perceived usefulness of the app or if the individual is feeling motivated to reduce their alcohol consumption.

State-Space Representation The state-space of an individual in a complete, complex dynamical system [97].

Just in time support Providing the right support at the right moment while avoiding the delivery of unhelpful support at other times.

Individualisation How information gathered about a person determines how the intervention will be delivered to that person.

Adaption A type of individualisation, where the sequence of multiple treatments are tailored to time-varying or dynamic information states of an individual over time. An example of a feeling state of engaging with *Drink Less* is feeling "lost and unsure of what to do next" with the app [59], as some users may feel disorientated within the app, and feel unsure how to navigate through it to achieve their goals[59].

Just-in-time Adaptive Intervention (JITAI) A dynamic treatment policy which aims to adapt the delivery of just-in-time support, to the individual's changing internal and external states over time. A JITAI is made up of six components: a distal outcome, a proximal outcome, decision points, tailoring variables and decision rules.

3.4.3 Components

Distal outcome The overall outcome the policy of the JITAI aims to improve, such as time to relapse of smoking, time to disengagement or

average number of steps taken.

Proximal outcome The short term goal of the intervention at that moment, such as steps taken 30 mins after receiving a notification, engagement with the app or completing self-report data.

Decision Points The point in time at which the intervention may be delivered.

Tailoring variables Time-varying or time-invariant information which is used about the individual that informs the adaption of the treatment policy.

Decision rules The operataionalisation of the JITAI, as a dynamic treatment policy, which specifies which intervention is offered for whom and under which states, current environment or past history. The decision rules can be informed by (i) exploring how a specific moment may alter the effect of a treatment on the proximal outcome, and (ii) the longer term effects of the sequencing of a treatment on the pathway to the distal outcome.

3.4.4 Example of a JITAI

An example of a JITAI is the A-CHESS app, which stands for Addiction-Comprehensive Health Education Support System [95, 42]. A-CHESS is a JITAI because it personalises the choice of treatment to the individual's

context through decision rules. One component of A-CHESS is that users specify locations they regularly purchase alcohol from. Through the use of global positioning systems, A-CHESS will send support to the user when they are near the prespecified locations, providing support.

3.5 Micro-Randomised Trials

Micro-Randomised Trial (MRT) is an experimental design to create Just-in-time Adaptive Intervention (JITAI) [222, 117].

In an MRT, individuals are randomised many times, perhaps hundreds or thousands of times, with known probabilities at pre-specified decision points. After each randomisation, a proximal, or near-term outcome is measured. When synchronised with contextual data, which may come from wearables, monitors or self-reports, data from an MRT is intensive, rich and longitudinal, made up of time-varying covariates, time-varying treatments and time-varying outcomes. The *history* of an individual is the evolution of such data for this person from the beginning of the trial. *States* can be captured by time-varying variable, or a summary of recent time-varying variables with time-invariant variables.

When developing "in the moment" interventions, researchers may naturally begin by understanding if there is, on average, any effect. If there is no marginal (i.e. average) effect, then any efforts to further optimise the inter-

vention, which does not have an overall effect, could arguably be futile.

Once the primary objective of understanding if there is a marginal effect, the repeat randomisation within individuals over time allows researchers to explore if the magnitude of the effect changes due to the measured, near-term recent states or there is an effect from the cumulative sequence of interventions over time. One example of a positive near-term effect but negative effect in the future (i.e long-term) is 'click-bait', which is designed to generate a near-term outcome, but over time is often a disappointing and unfulfilling experience, which can annoy users and reduce the perceived usefulness of the application in the long-term [165, 179].

The different types of effects which an MRT can estimate include:

Marginal effect Does the treatment, on average, work?

Treatment effect heterogeneity What works for one person may not work for other people

Contextual effect moderation What works for one user at one point in time may not work for the same person when they are in a different state

Short and long-term effect from sequencing of interventions The treatment may have a strong, near-term effect, but the cumulative sequence of treatments over time may create a negative effect on the ultimate

distal goal

Findings about such effects can help researchers develop better tailored interventions, which are more effective and less burdensome, while balancing near-term and long-term effects.

3.6 Conclusion

In this section I have described some research frameworks, trial designs and assessment methods which are key for the needs of developing and evaluating behaviour change apps or digital therapeutic apps.

The aims and objectives of the thesis are established in the next section.

Chapter 4

Aims and Objectives of the Thesis

Up to this point in the thesis, I have established the unique characteristics of behaviour change apps as health interventions. I have identified that apps can capture and adapt to an individual's changing state over time. When this personalisation is based on dynamic behaviour change theories, then behaviour change apps can deliver the right support to the right person at the right moment. That is, behaviour change apps can develop to become just-in-time adaptive interventions. I have also affirmed that engagement is considered an important mediator to optimise for effectiveness, with a conceptual framework which places behavioural engagement as a mediator and experiential engagement as a effect-moderator.

I have presented the current types of research methodologies to both capture outcomes, and how momentary interventions can treat individuals in their everyday lives, these are EMAs and EMIs. I have described the components of a JITAI, and how they can be developed from an MRT.

I now introduce *Drink Less*, an app which I base much of my research in this thesis with. Subsequently, I then set the aims and objectives of the thesis.

4.1 *Drink Less*

Drink Less is a theory- and evidence-based behaviour change app for adults from the general UK population who are seeking help to reduce their alcohol consumption [88]. The exact code of content for *Drink Less* is available on the Open Science Framework (https://osf.io/akqy9). This details how the app functions, such as the how the dashboard works, as well as details of how users can log their mood and drinks, how to set and change their goals and action plans.

The *Drink Less* modules include: (i) Normative Feedback, which consists of personalised feedback on how an individual's drinking behaviour compares with others' drinking; (ii) Goal Setting, which allows users to set weekly 'drinking reduction' goals, with brief advice on setting achievable goals; (iii) Cognitive bias retraining, delivered though a game which targets users' automatic biases through avoiding cues of alcoholic drinks and approaching

non-alcoholic drinks; (iv) Self-monitoring and Feedback, which allows users to monitor and reflect on their alcohol consumption, along with their mood, productivity, sleep and progress on goals; and (v) Action Planning, in which users create if-then plans for dealing with difficult drinking situations.

The development of *Drink Less* followed an iterative structure with feedback loops. The development strategy adopted the UK Medical Research Council guidance on complex interventions [57] and the Multiphase Optimisation Strategy [56], with the theoretical Capability, Opportunity, Motivation-Behavioural model driving the approach to module development within the complex intervention.

Two phases of development for *Drink Less* were undertaken. Phase One consisted of a scoping literature review, expert consensus study and content analysis of existing alcohol apps. Five components were selected from Phase One. These are (i) Normative Feedback, (ii) Cognitive Bias Re-training, (iii) Self-Monitoring and Feedback, (iv) Action Planning and (v) Identify Change.

Phase Two studied the acceptability and feasibility of the five components, with a person-based approach [234]. To understand the first impression of the app, a think-aloud study was conducted [59], with semi-structured interviews exploring users' impressions of longer-term use of the app in their natural settings.

Research into the refinement of *Drink Less*, within the optimisation phase,

included designing new app modules and content based on user feedback and improving the usability of the app through user testing [87]. The new modules include: Behavioural Substitution, Information about Antecedents and Insights.

The modules Normative Feedback and Self-Monitoring and Feedback both provide data collected by the individual over time. They are therapeutic components for individuals to track and monitor their own drinking habits over time and not designed to gather information for the device to learn about the individual over time and adapt the content. *Drink Less* is not an adaptive intervention, such as a JITAI but with further research can evolve to be one.

Drink Less is currently frozen (i.e. no incremental improvements are occurring) due to being evaluated in a parallel-group, conventional Randomised Controlled Trial [84]. The RCT aims to compare the effectiveness and cost-effectiveness of recommending *Drink Less* for reducing alcohol consumption to the National Health Service (NHS)) webpage on alcohol advice [84].

4.1.1 Funding of *Drink Less*

The development of *Drink Less* was funded by United Kingdom's Centre for Tobacco and Alcohol Studies, the National Institute for Health and Care Research (NIHR) School for Public Health Research and the Society for the Study of Addition and Cancer Research United Kingdom. The refinement

of *Drink Less* was funded by the National Institute for Health and Care Research (NIHR) School for Public Health Research and the Society for the Study of Addition [84] and the evaluation of *Drink Less* was funded by National Institute for Health and Care Research (NIHR) Public Health Research Programme [84].

4.2 Aims

The overall aim of this thesis is to improve engagement with *Drink Less*, with a particular focus on designing, conducting and reporting on a randomised trial to further optimise the notification policy.

4.3 Objectives

The stated aim will be achieved by focusing on the following objectives:

- Consider the types of evidence submitted to regulatory bodies, generated from randomised trials, for the approval of digital therapeutic apps, along with any challenges and limitation of such evidence.
- Identify novel trial designs to develop and evaluate behaviour change apps, and investigate their uptake for behaviour change apps.
- Explore the granular patterns of engagement with *Drink Less* over time.
- Design and conduct a Micro-Randomised Trial to understand and op-

timise the notification policy to improve engagement with *Drink Less*.

- Analyse the Micro-Randomised Trial and interpret the overall findings.

Chapter 5

Case Studies: A Tale of Two Apps

5.1 Current challenges of evaluating apps

Regulatory bodies overseeing the approval of digital therapies face novel challenges in a fast-moving territory. In this Chapter, I explore such challenges, through two case studies of digital therapeutic interventions which are both embedded with behaviour change components.

5.1.1 Case Study One

Reset by Pear Therapeutics evaluated by the Food and Drug Administration (FDA), federal agency of the United States Department of Health and Human Services.

Reset, developed by Pear Therapeutics, is a 12-week (90 day) FDA-cleared prescription digital therapeutic app to be used as an adjuvant to standard outpatient therapy for treating substance use disorder related to stimulants, cannabis, cocaine and alcohol [187]. In September 2017, the FDA permitted marketing of Reset as the first mobile medical application to help treat substance use disorders [129]. In accordance with the guidance document Medical Device Accessories – describing accessories and classification pathways, [152] the De Novo classification process provides a pathway for low to moderate risk accessories which are novel such that there are no legally marketed devices for comparison. The streamlined pathway requests special and general controls to provide a reasonable assurance of safety and effectiveness before clearance is granted. The media release for Reset states that, 'Prescription Digital Therapeutics are clinically-validated, FDA cleared software applications that demonstrate safety and efficacy in randomised clinical trials to improve patient outcomes.' The product is modelled on Community Reinforcement Approach [153], a form of cognitive behavioural therapy designed for patients with substance use disorder [105]. Pear Therapeutics' website states "reSET has been proven to increase abstinence from a pa-

tient's substances of abuse during treatment and increase patient retention in treatment when used as part of an outpatient treatment program". However no randomised controlled trial or other study has been published that evaluated the app Reset. Rather, a caveat on Pear Therapeutics' website explains "Therapeutic Education Systems (TES), which has equivalent content to reSET, was tested." The study submitted to the FDA to support Reset's effectiveness was a randomised controlled trial published by Campbell et al. in 2014 [39].

In the Campbell et al. study, patients seeking treatment for drug or alcohol problems were enrolled between June 2010 and August 2011. The trial evaluated TES, a computer-delivered intervention, made up of 62 web-based modules and contingency management components, to substitute about 2 hours of clinician time within treatment in a community centre. The contingency management component rewarded abstinence with voucher draws. Some of the draws result in a congratulatory "good job"; the others rewarded the participants with prizes of either small ($1), large ($20), or jumbo ($80–100) values in decreasing probability. During the 12-week intervention period, patients were assessed twice a week with urine tests for 10 different illicit drugs, and self-reported drug and alcohol use. The follow-up period, after the 12-week treatment programme ended, was at 3 months and 6 months. The average total of prizes per patient in the intervention arm was $277 over the 12-week treatment period. Campbell and co-authors reflect that, 'It may

be that the beneficial effect of TES observed during the active treatment phase was mainly attributable to the contingency management component of the intervention' (p. 689). That is, we are unable to disentangle the effect of TES with the effect of rewarding participants with prizes for abstinence.

The trial reported two primary outcome measures: abstinence from drug or heavy alcohol use in the last 4 weeks of treatment, and retention in treatment (time to drop out). A statistically significant treatment effect for abstinence was reported at the end of the 12-week treatment period, with an odds ratio of 1.62 (95% CI: 1.12, 2.35). However, at the 3- and 6-month follow-up points, 'The effect of TES compared with treatment as usual was no longer significant.' (p. 686).

This evidence submitted to the FDA for the approval of Reset raises questions about the content, usability, and efficacy of intervention. How does the content of TES, a computer-delivered version of community-based intervention, translate over to an app? How does delivery of the intervention through an app affect engagement, usability, and efficacy? Operational safety concerns can emerge when a device is transferred to an app, as discovered with Roche's Accu-Chek diabetes management app [231], which received its fifth FDA recall due to various software bugs leading to incorrect insulin dosage recommendations. Is the contingency management treatment a compulsory component of Reset, and if not, can the existing trial provide any relevant information regarding efficacy? If Reset is to be used as a single 12-week

therapy course, as described in the media press release, is there any evidence of long-term effectiveness?

5.1.2 Case Study Two

Deprexis by GAIA AG evaluated by National Institute for Health and Care Excellence (NICE) for National Health Service (NHS) programme Improving Access to Psychological Therapies for England.

The National Health Service (NHS) launched the Improving Access to Psychological Therapies (IAPT) programme in 2008, with the aim of expanding services to treat adult anxiety disorders and depression in England. In August 2017, NHS England and NICE announced up to 14 digital therapy products will be assessed for use of IAPT services by 2020 [218]. The call's eligibility requirements of digitally enabled therapy technologies, includes that "The content of treatment should mirror a NICE recommended psychological therapy" (p. 1); that it "should be designed to support a model of care where the therapist guides the user through the programme and regularly reviews the user's work, clinical outcomes and risk" (p. 1); "The technology must have at least one published randomised controlled trial" (p. 2); and it "must be supplied by an organisation committed to keep ownership and responsibility to maintain and update the technology." (p. 2). Deprexis is a cognitive behavioural therapy treatment for adults with depression, deliv-

ered through a mobile platform, which will be trialled within the NHS as part of NICE's evaluation for IAPT. The published literature on Deprexis includes eight randomised controlled trials [16, 23, 76, 118, 148, 22, 155, 32] four trial protocols [118, 147, 239, 120] and a systematic review [217]. The review [217], published in 2017, found that overall Deprexis was moderately effective.

Most of the Deprexis trials used a waitlist control, where participants in the control group receive access to the intervention after the initial 12-week intervention period, which prevented a thorough assessment of long-term effectiveness. The systematic review states that, "Using responsive design, Deprexis is optimized for use on any computer with internet access, including tablet computers and smartphones." However, the published trials do not mention use on a smartphone, so it is unclear how many trial participants were using it on a smartphone or how the optimisation for the mobile environment has been formally evaluated. Interestingly, Deprexis is an adaptive intervention which selects modules and content tailored to a patient's response, but there was no published literature on how such adaption, sequencing and decision rules were built. Therapists have access to the participants reported outcomes over time, can monitor how the patient is feeling and are alerted if someone's symptoms deteriorate, so Deprexis is in essence a dynamic treatment regime. As part of the IAPT approval process, Deprexis will first be adapted for integration within the NHS, and then evaluated in a

non-randomised trial (comparing with routinely collected data) of two-year duration. Whether the planned trial will assess the different elements of such a dynamic regime, as well as how successful the app is in collecting valid patient outcomes (perhaps through notifications acting as engagement prompts), remains unclear.

5.2 Implications for regulatory approval

In the already overcrowded app market, an agency's clearance will likely lead to a significant market advantage, but that clearance needs to be robust and fair. Regulatory agencies need to strike the right balance with the evidence required, which can include pre-approval clinical trials or post-approval surveillance, and this should be influenced by the app's potential clinical risk as a treatment, or the cost to society if it is widely adopted yet ineffective. The FDA is modernising its regulatory framework in line with recent 'Least Burdensome Provisions' [187, 46] to reduce 'outdated, unnecessary burden that can forestall beneficial innovation without also enhancing device safety and effectiveness'. There is an argument that requiring an evidence base only of numerous conventional randomised trials may not optimally enhance the safety and effectiveness of an app, and could forestall the roll out of a beneficial therapy to patients by a number of years. Plans for post approval maintenance of Reset in order to improve and sustain patient outcomes are unknown, nor do we have access to any details about how the ongoing eval-

uation of Deprexis will feed back into its design and optimisation of the intervention.

Apps often undergo continuous optimisation, learning from users and adjusting the treatment, but app developers may be unsure of how to develop clinically effective *adaptive* interventions in an evidence-based framework that allows for continuous, iterative optimisation. Without a way forward, we are at risk of using digital therapeutics in our health systems that are ineffective, unsafe, costly, and weaken evidence based clinical pathways. We require a fit-for-purpose framework where clinicians, statisticians and developers work together [150, 139] to support a timely and iterative process of knowledge transfer between developers, health care experts and end-users [131]. A necessary solution for such fit-for-purpose frameworks is that research and trial designs allow the development of dynamic, adaptive and data-rich interventions to be both effective interventions and transparent [222].

Chapter 6

Scoping Review

6.1 Trial designs for the development and evaluation of digital therapeutic apps: a scoping review

As discussed in previous chapters, for digital health to improve patients' lives, we must be able to easily distinguish efficacious treatments from products of commercial opportunism [126]. I explored how such challenges have impacted regulatory agencies in the past, through the Chapter *'A Tale of Two Apps'*. Regulatory bodies are working towards specific frameworks to facilitate such aims, yet challenges and shortcomings remain in these new frameworks as nimble guidelines can require either too little evidence for effectiveness and

safety, or long and costly evaluations can impede innovation and deter vital investment [178, 140, 3, 49]. Recently, high profile calls for better regulations of digital therapeutic apps have been made [13]. I have established that for an evaluation framework to succeed, fit-for-purpose clinical trial designs are required.

6.1.1 Aims of Scoping Review

This scoping review aims to bring together new trial designs which could suit the purpose of developing and evaluating digital therapeutic apps, and to explore results of their implementations.

As discussed in the previous chapters, the challenges of evaluating digital health interventions through conventional RCTs include long time frames which require the intervention to be 'frozen', and no longer leveraging the data gathered from individuals (from EMA or sensors/wearables) during the trial, to inform how the intervention can be continuously optimised. These limitations are discussed in the literature, which explored alternative methods, beyond the RCT, to evaluate digital health interventions [174, 150, 189, 66].

Previous research to explore alternative designs to the conventional RCT for the development or evaluation of digital health include three reviews. Law introduced adaptive trials of telehealth for more efficient research in 2014, and found no implementations of adaptive trials of telehealth in the literature

[128]. Pham's review in 2015 emphasised the poor fit of RCT for mHealth apps, due to rigid trial protocols, high implementation costs and long duration of recruitment, and her search of clinicaltrials.gov found no deviations from conventional randomised controlled trials for mHealth apps [174]. In 2018 McCallum considered how data from in-device sensors can assist rapid research designs, and concluded few activity apps and wearables were optimised for efficiency, engagement and acceptability [139]. In this chapter, I aim to update such previous reviews, and examine the implementation of novel trial methodology to develop and evaluate digital therapeutic apps.

6.1.2 Refining the research question

An initial literature search led to the following collection of three trial designs to under the MOST framework: (i) Sequential Multiple Assignment Randomised Trial (SMART) for optimising dynamic treatment regimens; (ii) Micro-Randomised Trials for developing Just-in-Time Adaptive Interventions; and (iii) factorial trials for optimising static complex interventions. I also included N-of-1 trials, as these are often trials undertaken in the field of personalised medicine. N-of-1 studies do not fall within the MOST framework for optimising a complex or dynamic intervention.

A brief explanation and example of each of these trial designs is provided in Appendix A.0.9.

Difference between optimised trials within the Multi-phase Optimisation Strategy Framework (MOST) framework and N-of-1 trials

I mapped out the common research objectives of a trial design in the context of a development phase in the MOST framework, or an evaluation trial. The development phase is the common iterative cycles of testing and review for digital products, also known as the optimisation phase, and the evaluation phase is set to determine effectiveness of the intervention in relation to a comparator.

Table 6.1: Trial design by intervention type, development or evaluation phase

Trials Design	Intervention Type	Development Phase	Evaluation Phase
Sequential Multiple Assignment Randomised Trials	Dynamic treatment regime	Yes	No, dynamic treatment regime evaluated in a RCT
Micro-randomisation trial	Just in time adaptive interventions	Yes	No, JITAI evaluated in a RCT
Factorial Trials	Complex Interventions	Yes	Yes, but within the MOST framework, likely to be an RCT
Series of N-of-1	Personalised treatments	No	Yes

"RCT" is shorthand for a traditional, 2-group, parallel arm trial.

6.1.3 Search Strategy for Identifying Novel Trial Designs

The searches were conducted the week beginning Monday 26th November 2018 with the database PubMed.

There is an umbrella of terms mobile heath interventions fall under. I analysed a range of relevant papers to identify applicable search terms. The resulting selection of search terms to identify interventions includes: mobile application; mobile health app; mobile health application; mobile app; mobile health intervention; smartphone application; smartphone app; web-based intervention; mobile health; mHealth; telemedicine; telehealth; eHealth; cellphone; handheld computer; user-interface and web-portal. These terms make up a broad definition of a mobile health intervention, to not miss any relevant papers.

The database search coordinated each trial design (Sequential Multiple Assignment Randomized Trials; Micro-Randomisation; N-of-1; Multiphase Optimization Strategy Framework) with all the mobile health search terms as stated above. Both American and British spelling was used for randomisation.

6.2 Objectives

The objectives of this scoping review are to (1) systematically search for implementations of the selected trial designs in mobile health; (2) consider the advantages and disadvantages of the trial designs; and (3) identify gaps of knowledge or previous limitations for future research in this area.

The review was drafted using the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [216]. In accordance with a framework for scoping reviews, this work followed the five phases (1) identifying the research questions; (2) identifying relevant studies; (3) study selection; (4) charting the data and (5) collating, summarising and reporting the results.

6.2.1 Key Research Questions

The following research questions, inclusion criteria and exclusion criteria were established.

1. What are the novel trial designs, alternative to the conventional, parallel-group randomised controlled trials, recommended in the literature for the development and evaluation of digital therapeutic apps?

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

3. What are the merits, challenges and future directions found from the

implementation of each trial design?

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

My goal is to select results papers from the selection of novel trial designs identified above. The intervention must aim to treat a particular health condition. The study design must have some mention of the designs selected (Sequential Multiple Assignment Randomized Trials; Micro-Randomisation; N-of-1; Multiphase Optimization Strategy Framework). The intervention must be an app to treat a health condition, as identified from the selection terms in the search strategy.

6.2.2 Inclusion criteria

- The app aims to treat and improve the user's health;
- The app is either a standalone intervention or part of a sequence or combination of therapies;
- The intervention can be delivered as an app on a smartphone, or may be available on a PC, laptop, tablet or website;
- The research paper is an original trial, feasibility or pilot study, protocol or statistical analysis plan;
- Trial outcomes were quantitative or qualitative;

- Trial outcomes were measures of efficacy, usability or engagement.
- The paper was published in English, indexed in PubMed, from any time point up to November 2018.

6.2.3 Exclusion criteria

- Apps which are for screening, diagnostics, data collection or communication purposes only;
- The app was only evaluated in a conventional parallel randomised controlled trial without any trials for the development phase;
- The app's intention is not to treat or improve a user's health;
- The research is methodological only, i.e, reports of simulations only;
- The paper is a viewpoint only.

6.2.4 Charting the data:

Both the CONSORT [197] and CONSORT-EHEALTH extension [73] were used to consider the advantages and disadvantages in the relevant studies selected. In particular, for the factorial trials under the MOST framework, I tabled the study aims, intervention, recruitment method and data collection. I also examined papers for quality assurance methods to ensure accuracy of outcomes and clarity of human involvement in the app's use and study

methodology. For all papers found, I report on and further discuss the outcomes, findings, limitations and future research raised.

Factorial Trials: Effect and Dummy coding

I will additionally consider the analysis of factorial trials, as discussed in a previous section of this thesis: *Effect coding or dummy coding*. This is because it is important to consider the different frameworks of ANOVA and regression for the analysis of data from factorial trials. How to analyse data from a factorial trial depends on the research question, and which estimators are appropriate. In this context of MOST, factorial trials are commonly undertaken in the optimisation phase for a complex intervention to determine the marginal effects of different therapeutic components. As discussed previously, when an interaction is fitted, effect coding is required to gain consistent estimates of all estimated effects. When applicable, I will also review if the protocol or trial results paper makes a mention of effect or dummy coding.

6.3 Results

6.3.1 Selection of sources of evidence

From the academic database Pubmed a total of 36 citations were found. The count of papers by trial design were Sequential Multiple Assignment Randomised Trials ($n = 4$); Micro-randomisation ($n = 6$); N-of-1 or Series of N-of-1 ($n = 15$); Multiphase Optimisation Strategy Framework ($n = 11$).

All papers' full text was assessed for eligibility. A flow diagram of the scoping view, and reasons for exclusion, by trial design are below.

6.3.2 Flow diagram of papers reviewed

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources

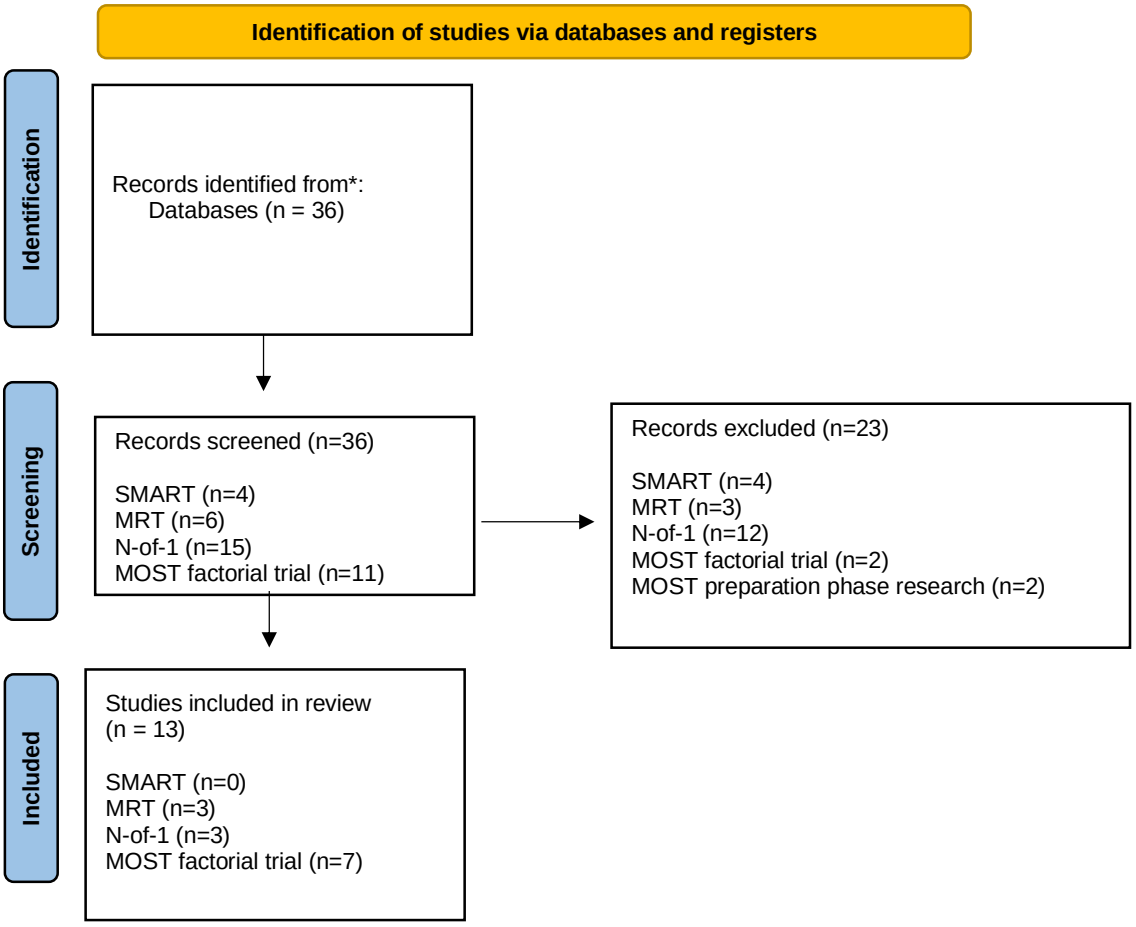


Table 6.2: Papers excluded by trial design, reasons for ineligibility, and number of papers.

Design	Reasons for ineligibility (n=number of papers).
Sequential Multiple Assignment Randomised Trials	Methodological only (n=4)
Micro-randomisation	Methodological only (n=3)
Factorial Trials under MOST	Methodological only (n=1) Systematic review (n=1) Preparation Phase (n=2)
N-of-1	Viewpoint or Letter to the Editor (n=6) Mobile device of data collection only (n=4) Systematic review (n=2)

Thirteen eligible studies from academic databases were found, and included Sequential Multiple Assignment Randomised Trials ($n = 0$); Micro-randomisation ($n = 3$); N-of-1 ($n = 3$), Multiphase Optimisation Strategy Framework ($n = 7$).

Some reasons for exclusion of the paper were (i) the app is an adaptive health intervention which personalises treatment over time, but only research known to be published was an evaluation in a conventional RCT [183] (ii) app as data collection device in the conduct of n-of-1 studies [12] and (iii) sequential multiple assignment trials to evaluate a sequence of SMS text messaging [20].

Characteristics of sources of evidence

I now tabulate the papers selected by trial design, and briefly describe the characteristics, objectives and findings of these studies selected. The tables

are split by trial design and the remarks regarding the papers found are provided by trial design, after each table is presented.

The Tables are

- Table 6.3 Summary of MRT research papers
- Table 6.4 Summary of N-of-1 research papers
- Table 6.5 Summary of factorial research papers
- Table 6.6 Study details of factorial research papers

Table 6.3: Summary of MRT research output papers

Author & Year	Title	Journal & Type of Paper
Kramer 2019	Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Protocol of a Microrandomized Trial	JMIR Research Protocols - Protocol
Rabbi 2018	Toward Increasing Engagement in Substance Use Data Collection: Development of the Substance Abuse Research Assistant App and Protocol for a Microrandomized Trial Using Adolescents and Emerging Adults	JMIR Research Protocols - Protocol
Bidargaddi 2018	To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App	JMIR Mhealth and Uhealth - Results Paper
Klasnja 2019	Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps	Ann Behva Med - Results Paper

Micro-Randomised Trial (MRT)

Four MRTs were identified from the search, as listed in Table 6.3.

For the MRTs, outcomes were efficacy orientated, with some engagement outcomes [116, 25, 185]. The largest sample size for a micro-randomised trial was 1,255 for engagement outcomes (user self-monitored behaviours and feelings 24 hours following a notification), the smallest was 44 for physical activity suggestions (step-count 30-minutes following a notification). No details of the validity or reliability of the data from wearables were provided. For all MRTs, the comparator treatment was no-notifications. None of the micro-randomisation trials included a power calculation to find a prespecified effect.

Klasnja et al. [116] provided a MRT for HeartSteps, a mHealth intervention which encourages regular walking though activity suggestions. The duration of the MRT was 6-weeks, during which participants were randomised five times a day. The primary outcome was 30-min step count following the decision time point of randomisation and the analysis employed a centred and weighted least squares method. The results found an increase in the 30-min step count, by 14% and the effect diluted over time. Future research directions discussed include understanding how well notifications work, for whom they work and do not work and what is the most optimal schedule.

Bidargaddi et al. [25] conducted a MRT to assess the causal effects of sending a push notification on proximal engagement for a wellness app. The decision

times were 1 of 6 times per day. The primary outcome was whether the user self-monitors behaviours and feelings sometime during the next 24-hours. The results showed risk ratio of 1.039 (95% CI 1.01 to 1.08), with effect moderation explored by weekends and over time.

The future discussion in Bidargaddi et al. [25] raises the following points:(i) treatment effect heterogeneity across users, as not all messages are likely to have the same effect for all users at all times, that (ii) sending messages when users are in receptive states could maximise engagement with mHealth apps, (iii) the burden and potential damage of self-reporting if implemented too frequently, and (iv) the promise for a more sensitive timescale measure, for the proximal outcome (as anything finer than a 24-hour period was not available from this study).

Rabbi et al. [185] published an MRT protocol with SARA, a mobile application to increase or sustain engagement of substance data collection over time. The study recruited participants who were aged 14-24 years old and asked to use SARA for 1 month. The primary analysis was to determine if sending a push notification at 4pm increased self-reporting on the current or following day. In Rabbi's [185] protocol paper, the discussion section mentions the potential for initialising machine learning algorithms, with the aim to develop engagement strategies which adapt to the personal context of users. The benefit of the SARA study [185] is the rich nature of the data gathered over time though the mobile data collection app, which allows for

future research to study in-the-moment precedents and sequelae of substance use among adolescents and young adults. An understanding of such temporal patterns of substance abuse in the individual's context is required to inform the development of interventions.

Kramer et al [119] is a protocol for an MRT, for the aim of quantifying effects (main, interactions and moderators) of three components in an app to promote physical activity. The MRT was 6-week period, and users were randomised daily to either receive a prompt to encourage weekly self-reporting. The proximal outcome is the proportion of overall participant days that the step goal is achieved during the intervention period. The analysis would use a centred and weight least squares model, to avoid biased effects from the alternative GEE-based approach.

Table 6.4: Summary of N-of-1 research output papers

Author & Year	Title	Journal & Type of Paper
Brannon 2018	Goal feedback from whom? A physical activity intervention using an N-of-1 RCT	Psychology & Health - Results Paper
Quinn 2013	Testing an integrated behavioural and biomedical model of disability in N-of-1 studies with chronic pain	Psychology & Health - Results Paper
Yoon 2018	Using Behavioral Analytics to Increase Exercise: A Randomized N-of-1 Study	American Journal of Preventive Medicine - Results Paper

N-of-1 Trials

Three N-of-1 Trials were identified: Brannon et al. [30], Quinn et al. [182] and Yoon et al. [235].

Brannon et al. [30] undertook an aggregated N-of-1 study, to answer the research question *for whom did the intervention work?*. The intervention was a SMS text message to increase physical activity. The message was sent from a parent, a peer or a behavioural specialist. Outcomes were measured through MyFitnessPal and include heart rate, moderate and vigorous physical activity through steps taken and calorie counting. The results did not report on marginal effects, with only significant interaction effect or with-persons effects reported on. The results focus on treatment effect heterogeneity, with some adolescents responding to feedback on goal attainment to increase physical activity. The discussion proposes that future research should aim to better understand the contextual factors which modified the effect of the text message on physical activity outcomes.

Quinn et al. [182] published the results of an N-of-1 trial in 2013. In this study, six women with arthritis and walking limitations participated in the study, an were requested twice daily to self-report measures for 60-90 days. No intervention was applied. The goal of the study was to examine theoretical models of both behavioral and biomedical within-person level. The paper raises the issues described as autocorrelation or serial dependency over time, and how previous similar studies used structural equation models to account

for this. The outcomes were collected by hand-held computers.

Yoon et al. [235] studied the impact of behavioral interventions of physical activity. The study design is a 12-month observational study design which collected intensive longitudinal data through an accelerometry (Fitbit Flex). Participants were randomised once, at the 6-month mark into the study, to either receive a tailored message about their personal predictors of exercise or not. The creation of the personalised message was informed by the 6-months prior accelerometry data.

Table 6.5: Summary of factorial trials under the research output papers

Author & Year	Title	Type of Paper	Journal	Study Design
Burman 2016	BeWell24: development and process evaluation of a smartphone 'app' to improve sleep, sedentary and active behaviors in US Veterans with increased metabolic risk	Original Research Evaluation of an app	TBM	full factorial 2 x 2 x 2 screening experiment
Kugler 2016	Application of the multiphase optimization strategy to a pilot study: an empirical example targeting obesity among children of low-income mothers	Technical Advance	BMC Public Health	fractional factorial design
Pellegrini 2014	Optimisation of remotely delivered Intensive Lifestyle Treatment for Obesity using MOST: opt in study protocol	Study Protocol	Contemporary Clinical Trials	fractional factorial design

Table 6.5: Summary of factorial trials under the research output papers

Author & Year	Title	Type of Paper	Journal	Study Design
Phillips 2018	Optimisation of a technology-supported physical activity intervention for breast cancer survivors: Fit2Thrive study protocol	Study Protocol	Contemporary Clinical Trials	full factorial design
Crane 2018	A smartphone to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial	Original Research	Scientific Reports	factorial trial

Table 6.5: Summary of factorial trials under the research output papers

Author & Year	Title	Type of Paper	Journal	Study Design
Uwatoko 2018	Healthy Campus Trial: a multiphase optimization strategy (MOST) fully factorial trial to optimize the smartphone cognitive behavioral therapy (CBT) app for mental health promotion among university students: study protocol for a randomized controlled trial	Study Protocol	Trials	full factorial trial
McClure 2013	The effect of program design on engagement with an internet-based smoking intervention	Original Research	J Med Internet Res	2-level full factorial trial

Table 6.5: Summary of factorial trials under the research output papers

Author & Year	Title	Type of Paper	Journal	Study Design
Tombor 2016	Development of SmokeFree Baby: a smoking cessation smartphone app for pregnant smokers	Original Research	TBM	Preparation phase: review relevant scientific literature
Garnett 2018	The development of Drink Less; an alcohol reduction smartphone app for excessive drinkers	Original Research	TBM	scoping literature review & expert consensus study

Table 6.6: Study details of MOST papers

Author & Year	Phase of MOST & Study aims	Intervention	Recruitment methods	Data collection
Buman 2016	Optimisation Phase: 8-week multiphase optimisation strategy design was used to test the initial efficacy of the sleep, sedentary and exercise components of the app	BeWell24, a multicomponent app targets behaviour change that make up the 24h spectrum (sleep, sedentary behavior and physical activity)	distributing flyers in clinic waiting and exam rooms, targeted mailings	Usage data and qualitative interviews

Table 6.6: Study details of MOST papers

Author & Year	Phase of MOST & Study aims	Intervention	Recruitment methods	Data collection
Kugler 2016	Optimisation Phase: Pilot study to identify the most feasible and acceptable components for low-income mothers with and without depressive symptoms and their 12 to 42 month old children	a remotely delivered responsive parenting intervention	Recruitment from WIC clinics	Telephone, semi-structured interviews

Table 6.6: Study details of MOST papers

Author & Year	Phase of MOST & Study aims	Intervention	Recruitment methods	Data collection
Pellegrini 2014	Optimisation Phase: to dentify which of 5 treatment components or component levels contribute most meaningfully and cost-efficiently to the improvement of weight loss over a 6 month period	weight loss treatment	remotely delivered intensive lifestyle treatment for obesity	Online posting and advertisements, telephone screening semi-structured interviews

Table 6.6: Study details of MOST papers

Author & Year	Phase of MOST & Study aims	Intervention	Recruitment methods	Data collection
Phillips 2018	Optimisation Phase: Identify which of 5 potential technology-supported intervention components would contribute to increased MVPA among breast cancer survivors	Core intervention which includes a Fitbit and standard self-monitoring Fit2Thrive smartphone application	All online or over the phone	ActiGraph GT3X-BT, a valid and reliable physical activity measure
Crane 2018	Optimisation Phase: Evaluate 5 intervention components of an alcohol reduction app	Behaviour Change App	Online	Online

Table 6.6: Study details of MOST papers

Author & Year	Phase of MOST & Study aims	Intervention	Recruitment methods	Data collection
Uwatoko 2018	Optimisation Phase: Select components for optimize an app for mental health promotion among university students	smartphone cognitive behavioral therapy (CBT) app	In person	Online Collection Measures
McClure 2013	Optimisation Phase: Select from 4 components which features increase treatment engagement	an online smoking cessation program	online	online engagement measures

Table 6.6: Study details of MOST papers

Author & Year	Phase of MOST & Study aims	Intervention	Recruitment methods	Data collection
Tombor 2016	Preparation Phase : Review relevant scientific literature	smartphone app designed to aid smoking cessation during pregnancy	NA	NA
Garnett 2018	Preparation Phase: Scoping literature review, expert consensus study and content analysis of existing alcohol apps	Behaviour Change App	NA	NA

Factorial Trials

The eligible trials for the factorial design include seven studies. Four papers were results [141, 58, 34, 121], three protocols [170, 177, 219]. Two papers discuss the *preparation phase* research which informed the components of the complex intervention and design of the factorial trial [88, 213].

The range of factorial trials include qualitative outcomes, pilot studies with fractional factorial design, powered trials without interactions, but with interactions explored and powered trials for an interaction [58, 34, 121, 177].

Results papers

Four papers [34, 121, 59, 141] were results papers from factorial trials to optimise a complex intervention.

Buman [34] performed a full factorial trial for the development and process evaluation of a smartphone app to improve sleep, sedentary and active behaviors in US Veterans with increased metabolic risk. The intervention is BeWell, for US Veterans currently receiving clinical care at a regional VHA hospital in the SouthWestern United States, aged 35-60, measured as overweight/obese, fasting glucose over 100 mg/dL. The trial was 8-weeks with a sample size of 26 participants. The process evaluation outcomes included app usage (minutes of usage, self-monitoring patterns), and this data was analysed by graphical displays over time. The factorial trial collected qualitative outcomes.

Kugler [121] published a pilot study, undertaken in the framework of MOST, to optimise an intervention to treat obesity among children of low-income mothers. The study was a fractional factorial trial, with the objective to assess the acceptability of the intervention. The outcome was the overall completion rates of the intervention and the number of times that research staff attempted to contact a participant via phone or text message. The findings include mothers with depressive symptoms had lower completion rates than mothers without depressive symptoms. No mention of effect coding was made.

Crane [58] implemented a factorial trial to develop *Drink Less*, an alcohol reduction app, and evaluate intervention components. The primary outcome was a change in weekly alcohol units consumed with a sample size 672. The analysis was a one-way ANOVA, however no mention of effect coding was made, it is possible effect coding was implemented through the ANOVA analysis.

McClure [141] is the earliest results paper, published in 2013. This research examined which intervention modules increased engagement in a online smoking cessation programme. The design was a 2-level, full factorial design testing the interventions message tone (prescriptive vs motivational), navigation autonomy (dictated vs not), proactive email reminders (yes vs no), and inclusion of personally tailored testimonials (yes vs no). The outcomes were number of visits to the website resulting in intervention content views (as op-

posed to supplemental content views), number of intervention content areas viewed, number of intervention content pages viewed, and duration of time spent viewing this content. The results report that users receiving proactive email reminders made 1.20 times as many visits (95% CI 1.09-1.33).

Protocols

Three papers [170, 176, 219] were protocols using factorial trials to optimise a complex intervention.

Pellegrini [170] published a protocol to optimise a remotely delivered treatment for obesity. A fractional factorial trial was designed, with the binary outcome of weight loss greater than 7%. Recruitment was all in-person. A linear mixed model was stated in the analysis plan, to test whether each factor has a significant effect on weight change across the time points (baseline, 3- & 6-months). No mention of effect or dummy coding was made.

Phillips [177] also published a protocol for a full factorial trial to optimise Fit2Thrive, a physical activity intervention for breast cancer survivors. Five interventions were tested, and the primary outcome was average daily minutes of moderate to vigorous physical activity (MVPA) as measured by accelerometry. The study recruited participants and collected outcomes all online. A generalised linear model was used to estimate the change in MVPA over time. Effect coding was stated and described in the protocol.

Uwatoko [219] designed a fully factorial trial to optimize the smartphone

cognitive behavioral therapy (CBT) app for mental health promotion among university students. Five components were tested, with in-person enrollment. The primary outcome is a change in PHQ-9 scores, and effect coding was detailed in the analysis plan, as experimental factors will be coded at two levels (presence coded as +1, and absence coded as -1).

6.3.3 Research papers excluded from the scoping review but of interest

Two papers were not factorial trials, but undertook research to inform the factorial trial for the optimisation phase [213, 88] in the MOST framework. I discuss this research as it relates to the theoretical frameworks and research to develop behaviour change apps.

Tombor [213] was guided by the Behaviour Change Wheel and MOST framework to develop an app to aid smoking cessation during pregnancy. The research included a identification of a theoretical base as the COM-B and a systematic review to select intervention components. The five modules selected are (i) identify, (ii) stress relief, (iii) health effects (iv) face-to-face and (v) behavioural substitution. The targeted behaviours include psychological capability, social opportunity, environmental opportunity, automatic motivation and reflective motivation. The modules were further prepared for the next phase of MOST, as the factorial trial for screening purposes.

Garnett [88] under took a scoping literature review, expert consensus study

and content analysis to select the modules i) Normative Feedback, (ii) Cognitive Bias Re-training, (iii) Self-monitoring and Feedback, (iv) Action Planning, and (v) Identity Change for the development of *Drink Less*.

N-of-1 trials with apps for data collection only

I will briefly discuss the characteristics of two N-of-1 trials that were excluded from the results table, as the purpose of the app was for facilitating the conduct of a N-of-1 trial and/or the collection of outcomes for during the N-of-1 trial. This is to demonstrate how N-of-1 trials can be conducted through an app, but do not include or intend to include an therapeutic interventions. The PREEMPT trial [12], a two arm RCT, randomised patients to participate either the arm (A) N-of-1 trial, with treatment reminders and data collection through the app *Trialist*, to help develop a individualised treatment plan for patients with chronic pain, or the arm (B) allocated to usual care. The PREEMPT trial is similar to the StatinWISE trial [100], a series of N-of-1 trials to determine if adverse muscle effects are caused by statin use at both an individual and marginal level. This trial allowed participants to report their data collection through a bespoke mobile app on their own smartphones.

6.3.4 Adaptive interventions for personalisation and adaptive trials for efficiency

Some MRT result papers mention future steps of initialising machine learning algorithms to optimise the intervention, to deliver more personalised content [184, 25]. This future work, which builds on MRT results, may look to algorithms such as the multi-arm bandit model or Thompson sampling [63]. Thompson Sampling is based Bayesian methodology which learns from the data collected over time, with the randomisation of the next treatment selected conditional on the posterior probability of that treatment being the most optimal at that time [63].

In the clinical trial context, response adaptive trials are undertaken with the goal to design more efficient or ethical trials. That is, to obtain robust clinical findings in a quicker manner with fewer patients randomised to a less efficient treatment [221, 168]. However, in the context of MRTs, the adaptive algorithm is part of the development of the adaptive intervention.

6.4 Conclusions

Principal Findings

For the development of behaviour change apps, which aim to treat chronic conditions over time, with some behavioral components, there is an emerging uptake of micro-randomised trials and factorial studies (both preparation

stage and optimisation stage). At this point in time (late 2018), no SMART trials were identified for the development of digital therapeutic apps. Two N-of-1 studies made some mention of the objectives to personalise the intervention to the contextual circumstances of the intervention. A common challenge among all on-line trials is low follow-up rates.

Just-in-time adaptive interventions allow for the sequencing and adaption within an intervention for the fast-paced contextual circumstances that naturally arise through mobile health, whereas SMARTs are designed for a more slow-paced sequencing and adaption in dynamic treatment regimes [44] found in a clinical setting [159]. As mobile health interventions become more integrated into clinical care, such as *Deprexis* (as described in *A Tale of Two Apps*, where *Deprexis* has the first stage intervention of using the app, and if the patient is not responding to this as a treatment, a second stage is consultation with the GP begins), then more hybird designs between MRTs and SMARTs will be more suited [161].

MRTs

The importance of optimising engagement as well as effectiveness is emerging. Rabbi leads with a two phase MRT process [185, 184]. The first MRT tests various engagement and follow-up components, then when sufficient engagement and follow-up is established, Rabbi mentions a second MRT may test therapeutic interventions. This approach mitigates the universal risk of

testing a multi-component app for effectiveness in a trial, only to discover engagement for self-reporting outcomes was too poor to tailor the treatment and understand any attributable effectiveness. More options and guidance on how researchers can consider and balance the aims of optimising engagement and effectiveness components could be helpful.

The two most helpful papers were the JOOL MRT [25] and the SARA [184], as I became aware of the importance of burden and potential damage to longer-term engagement if the randomisation policy implemented in an MRT is too frequent, the gap in knowledge of a more immediate, finer proximal outcome in time and understanding the theoretical links and synergies between engagement and effectiveness as a temporal process within a dynamic individual environment, to achieve the overall objective defined as the distal goal.

Factorial trials for optimisation objectives

The review picked up on two papers [85, 213] under the MOST search, which reported on earlier work in the *preparation phase*, to consider theoretical frameworks to inform the development of interventions in the *optimisation* phase. Although these are not randomised trials, I have kept the papers in as they are important components of the MOST framework.

I found that effect coding was mentioned in most protocols which implemented a factorial trial for the optimisation phase of the MOST framework.

The number of module components evaluated was commonly five. Interaction effects were estimated for various components and it is possible that the inflation of type I error rates may need to be considered.

N-of-1

N-of-1 studies involve repeated cycles of treatments and outcome measures over time. The early study of Quinn [182], published in 2013, had the objectives of learning about individual characteristics to create a tailored, personalised intervention in the future. These objective can be achieved through the framework of MRTs. The role of mobile apps to collect this data during the N-of-1 and MRTs is emerging. This is akin to Ecological Momentary Assessment (EMA) studies, and a shared understanding of the similarities and differences between N-of-1 studies, MRTs and Ecological Momentary Assessment (EMA) may be helpful to the community who are pursuing the development of personalised medicine through apps.

What are the opportunities for improving trial methodology for digital therapeutic apps?

As discussed in previous chapters, trial designs which synchronise and balance the two separate goals of learning how to continually improve an intervention, while gathering information to assess overall effectiveness, would be helpful in the realm of digital health. This review, undertaken in 2018, revealed an emerging uptake of MRTs and factorial trials for the development of

behaviour change apps. These novel trial designs and frameworks are fit-for-purpose to meet this need. A potential opportunity to increase research quality includes the development of formal guidance, such as a CONSORT statement, for reporting such trial designs may be helpful.

When apps are currently used in the real-world, it is common for app developers to continuously improve the intervention through iterative cycles of development. In terms of developing complex apps, the rapid accumulation of real-time data from observational studies lends itself to Bayesian methods, such as historical controls, for the optimisation of an intervention. This may be helpful when information to contextually tailor the intervention can be sourced from both cohorts studies, qualitative studies and ongoing trials.

To conclude this section of the thesis, previous reviews by Law (2014) [128], Pham (2016) [174] and McCallum (2018) [139] found no evidence of any uptake of novel trial designs which suit the needs for developing or evaluating multi-component behaviour change apps. Whereas this review, undertaken in 2018, found the implementation of such trial designs emerging.

Prior to 2015, research teams aimed to tailor the sequence of treatment to the individual's context and prior history. MRTs now offer a framework to personalise the sequencing of individual treatments over time.

I now move on to my research with *Drink Less*, exercising my knowledge gained up to this process, about MRTs, the importance of engagement and

setting up the optimisation of a static behaviour change app to evolve into a JITAI.

Chapter 7

Paper One. Engagement with a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study

From the systematic review into trial designs for the development and evaluation of behaviour change apps, I concluded that engagement was sometimes the primary outcome of such developmental trials, and that there is an emerging uptake of JITAIs to improve engagement or self-reporting, and

such research was undertaken through an MRT.

Going forward, I take inspiration from two particular MRTs which focus on engagement: the SARA MRT [184] and the JOOL MRT [25].

Engagement is a multifaceted construct, which comprises behavioural and experiential aspects [171], and usage data offers researchers and app developers an opportunity to explore how the app is used over time, by the help-seeking individuals who download the app. There is a growing appreciation for understanding the complex engagement patterns within and between users over time [175, 157].

Knowledge about how people actually use the app can offer insights to (1) consider how such patterns of use may boost or hinder behaviour change, and (2) consider pathways to further incrementally improve the intervention [122]. The usage data from behaviour change apps is voluminous and temporally dense. App developers and researchers may find it challenging to easily summarise such rich and big data, to provide any usefully insights.

In this Chapter, I begin by exploring the engagement data from large cohort of users (n=19,233) with *Drink Less* though simple, accessible visualisations, primarily with the R package *ggplot* [225]. I demonstrate that such simple visualisations offer valuable insight into the patterns of use over time.

In reflection, this work also showed that current notification is likely to affect engagement more than I expected, and that the fixed notification policy was

effective for maintaining engagement for some users, but other users may habituate to the daily notification.

From the insight gained from the simple visualisation, I concluded that optimising the notification was a worthwhile investment. This conclusion paved the way forward for a future MRT.

SCHOOL of HYGIENE &TROPICAL MEDICINE

Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1703883	**Title**	Ms
First Name(s)	Lauren Marie		
Surname/Family Name	Bell		
Thesis Title	Designing Randomised Trials to Improve Engagement through Optimising the Notification Policy of a Behaviour Change App		
Primary Supervisor	Professor Elizabeth Williamson		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Journal of Medicial Internet Research		
When was the work published?	11th December 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	119
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	LB conceptualised and designed the study with contributions from CG, TQ, OP, HP and EW. CG extracted the data. LB manipulated and analysed the data with key support from HP and EW. LB prepared the manuscript and wrote the first draft. All authors contributed to and approved the final version of the published paper. LB responded to peer review comments with support from all authors.

SECTION E

Student Signature	Lauren Bell
Date	20/06/2022

Supervisor Signature	Elizabeth Williamson
Date	20/06/2022

"The greatest value of a picture
is when it forces us to notice
what we never expected to see."

— John Wilder Tukey, 1977,
Exploratory Data Analysis

Original Paper

Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study

Lauren Bell[1], MSc; Claire Garnett[2], PhD; Tianchen Qian[3], PhD; Olga Perski[2], PhD; Elizabeth Williamson[1,4], PhD; Henry WW Potts[4,5], PhD

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2]Research Department of Behavioural Science and Health, University College London, London, United Kingdom

[3]Department of Statistics, University of California Irvine, Irvine, CA, United States

[4]Health Data Research UK, London, United Kingdom

[5]Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Lauren Bell, MSc
Department of Medical Statistics
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 02076368636
Email: lauren.bell@lshtm.ac.uk

Abstract

Background: Behavior change apps can develop iteratively, where the app evolves into a complex, dynamic, or personalized intervention through cycles of research, development, and implementation. Understanding how existing users engage with an app (eg, frequency, amount, depth, and duration of use) can help guide further incremental improvements. We aim to explore how simple visualizations can provide a good understanding of temporal patterns of engagement, as usage data are often longitudinal and rich.

Objective: This study aims to visualize behavioral engagement with *Drink Less*, a behavior change app to help reduce hazardous and harmful alcohol consumption in the general adult population of the United Kingdom.

Methods: We explored behavioral engagement among 19,233 existing users of *Drink Less*. Users were included in the sample if they were from the United Kingdom; were 18 years or older; were interested in reducing their alcohol consumption; had a baseline Alcohol Use Disorders Identification Test score of 8 or above, indicative of excessive drinking; and had downloaded the app between May 17, 2017, and January 22, 2019 (615 days). Measures of when sessions begin, length of sessions, time to disengagement, and patterns of use were visualized with heat maps, timeline plots, k-modes clustering analyses, and Kaplan-Meier plots.

Results: The daily 11 AM notification is strongly associated with a change in engagement in the following hour; reduction in behavioral engagement over time, with 50.00% (9617/19,233) of users disengaging (defined as no use for 7 or more consecutive days) 22 days after download; identification of 3 distinct trajectories of use, namely engagers (4651/19,233, 24.18% of users), slow disengagers (3679/19,233, 19.13% of users), and fast disengagers (10,903/19,233, 56.68% of users); and limited depth of engagement with 85.076% (7,095,348/8,340,005) of screen views occurring within the *Self-monitoring and Feedback* module. In addition, a peak of both frequency and amount of time spent per session was observed in the evenings.

Conclusions: Visualizations play an important role in understanding engagement with behavior change apps. Here, we discuss how simple visualizations helped identify important patterns of engagement with *Drink Less*. Our visualizations of behavioral engagement suggest that the daily notification substantially impacts engagement. Furthermore, the visualizations suggest that a fixed notification policy can be effective for maintaining engagement for some users but ineffective for others. We conclude that optimizing the notification policy to target both effectiveness and engagement is a worthwhile investment. Our future goal is to both understand the causal effect of the notification on engagement and further optimize the notification policy within *Drink Less* by tailoring to contextual circumstances of individuals over time. Such tailoring will be informed from the findings of our

micro-randomized trial (MRT), and these visualizations were useful in both gaining a better understanding of engagement and designing the MRT.

(J Med Internet Res 2020;22(12):e23369) doi: 10.2196/23369

KEYWORDS

mobile health; behavior change; apps; digital health; data visualizations; engagement; micro-randomized trial; push notifications; just-in-time adaptive interventions

Introduction

Background

Maintaining alcohol consumption within recommended guidance is widely known to reduce one's risk of illness or injuries. Such guidance includes the recommendations of the Chief Medical Officer of the United Kingdom to limit alcohol consumption to 14 units a week and to have frequent alcohol-free days [1]. However, anyone in the general adult population who wants to reduce their hazardous or harmful alcohol consumption may face certain challenges to follow such guidance [2]. Challenges include the ease of access to alcohol and alcohol being an addictive substance. This can lead to individuals developing chronic or cyclical patterns of excessive drinking, with the personal behaviors of drinking influenced by internal or external factors [3-5]. Internal factors refer to feeling states or events in an individual's recent drinking history, such as previous drinking episodes, moods, motives, or cravings that may modify future patterns of drinking [6]. External factors are influential events that occur independently of an individual's drinking history; for example, how the risk of hazardous drinking of the general population increases during holiday periods or weekends [7,8].

Behavior change apps, sensors, and wearables offer a way of reducing hazardous alcohol consumption through real-time data capture and interventions [9-12]. Benefits of behavior change apps, that can be synchronized with sensors and wearables, include capturing an individual's dynamic history of alcohol consumption and state of mind while providing *around the clock* access to support, particularly in moments when an individual's vulnerability to hazardous drinking may increase [13].

However, a key challenge for the majority of behavior change apps is that levels of engagement remain low [14-16]. Engagement, often a mediator of effectiveness [14], is considered a multifaceted construct composed of behavioral and experiential aspects [17]. Usage data from a behavior change app provides an understanding of behavioral engagement (hereafter referred to as engagement) with the app [18]. Multiple indicators of engagement are thought to convey important information about how users interact with a given intervention, including the frequency (eg, number of log-ins), depth (eg, proportion of available modules accessed), amount (eg, time spent per log-in), and duration (eg, total number of days) of use [19].

Drink Less is a behavior change app that aims to help its users reduce hazardous and harmful alcohol consumption. The app was developed following the multiphase optimization strategy framework (comprising a preparation phase, an optimization phase, and an evaluation phase) [20-23] and the UK Medical Research Council's guidance on developing complex interventions [24-26]. The app includes 6 different theory and evidence-informed modules: normative feedback, goal setting, cognitive bias training, self-monitoring and feedback, action planning, and identity change. These modules are described in detail by Garnett et al [27]. The app sends a local daily push notification at 11 AM that asks users to "Please complete your drink diaries," to encourage self-monitoring of drinking behavior. The default 11 AM timepoint was set so as not to disturb late risers and to allow participants time to complete their morning routine; however, the notification timing could be changed by the user.

Owing to the agile nature of app development, optimization of engagement can be done through cycles of research and implementation [28]. Identifying important patterns of engagement for such optimization purposes presents various analytical challenges that visualizations can address. Visualizations have previously been helpful for analyzing a wide variety of rich data streams within public health research [29-33]. Simple visualizations, especially when complemented with clear textual descriptions, are generally recommended for identifying and comparing trends [32]. In previous digital health research, visualizations have delivered *at a glance* insight from mass volume and time-varying data, including more sophisticated displays of spatiotemporal, contextual, and event-centric outcomes [34-38]. Importantly, visualizations can provide insights into optimization that include (1) patterns of use that may boost or hinder behavior change, (2) a better understanding of temporal engagement with various components of the intervention, and (3) pathways toward personalization of the intervention.

Objectives

The aim of this paper is to explore the usefulness of simple visualizations in uncovering important temporal patterns of engagement and facilitating decision making for further intervention development. This study presents 2 key contributions to improving engagement with *Drink Less*. The first contribution, provided in the Results section, is to showcase a number of visualizations that helped us understand temporal patterns of engagement with *Drink Less*. The second contribution, provided in the Discussion section, explains how insights obtained from these visualizations informed the next stages of intervention optimization.

123

Methods

Data Transformation

Each visualization involved transformation of the data. Original usage data involved merging, by an anonymous user ID, a data

set of baseline characteristics (age, sex, employment type, and Alcohol Use Disorders Identification Test [AUDIT] score) to a data set of time stamps of start time of use, screen views, and length (in microseconds) of use. Along with use, the actions of entering an alcohol-free day or recording units of alcohol consumed were measured.

Data

Data set 1 included 19,233 users who downloaded *Drink Less* between May 17, 2017, and January 22, 2019 (615 days). The inclusion criteria for users included having a baseline AUDIT score of 8 or above, which is indicative of excessive drinking [39]; being from the United Kingdom; being aged 18 years or above; being interested in reducing their alcohol consumption; using app versions 1.0.11 to 1.0.16; and having consented to the Privacy Notice (Multimedia Appendix 1). Screen views data are recorded automatically and downloaded via Panda scripts from *Nodechef* (a web-based platform for hosting mobile apps) using a secure https protocol. Sessions were derived from screen views using the Pandas script.

Users who downloaded the app on August 21, 2018 (n=5830), were excluded as an article on BBC News was published on this date, which endorsed the app (Garnett et al, unpublished data, 2020); thus, these users were likely to have different characteristics and engagement behavior.

Data set 2 included time stamps of 829,001 sessions and 8,169,005 screen views of the 19,233 users in data set 1. This includes 122,332 entries of alcohol-free days and 123,704 entries of alcohol drinks consumed. All use was recorded from May 17, 2017, to April 16, 2019 (699 days). As such, users had a minimum of 84 days of use measured.

To explore various engagement aspects, we developed sets of data from data sets 1 and 2 with varying engagement measures.

Set A

All use was measured from May 17, 2017, to January 22, 2019 (615 days), including date of download and time stamps of all use. This period was chosen as it reflects a time in which the content of the app was relatively stable.

Set B

Set B included all users whose use was measured in Set A, with data only over the first 30 days from download, with the measure "Did use occur on this day?" (binary, yes or no) for each user.

Measures

Log-in Sessions and Frequency of Log-Ins

A session was defined as a continuous series of screen views, with a new session defined as a new screen view after 30 min of inactivity [40]. Clearing or *swiping away* the daily notification did not register as use and was not considered as either a session or a module view. All time stamps were appropriately adjusted from Coordinated Universal Time to British Summer Time. The amount of use per log-in session was operationalized as time spent (in seconds) per session. Daily use was captured by the measure "Did use occur on this day?" (binary, yes or no) for each user for 30 days (Set B).

Drinking Diary Entry

In the self-monitoring and feedback module, users enter an alcohol-free day and the date of its occurrence, the number of alcoholic units consumed, and the date of consumption. The time stamps in which records were made was measured.

Disengagement

We defined disengagement as the first day of 7 or more consecutive days of no use after download [41]. The days between download and disengagement were derived for each user. Users who did not disengage after downloading the app were censored.

Data Visualization Methods and Analytical Techniques

We used heat maps, timeline plots, k-modes clustering, generalized estimation equations, and Kaplan-Meier plots to explore and visualize patterns of engagement with *Drink Less*. Analyses were carried out in R [42] and Stata [43]. We used the following R library packages to create the visualizations: ggplot2 for heat maps and timeline plots [44], rayshader to create the 3D animations [45], viridis for color palettes sensitive to readers with color blindness [46], Klar to perform the k-modes clustering [47], survminer for the Kaplan-Meier survival curves and number at risk table [48], gganimate to create animated plots of use over time [49], and patchwork to place graphs side by side [50]. The data visualization methods, data set and engagement measures are shown in Textbox 1.

K-modes clustering is an extension of the k-means algorithm for partitioning categorical data, which uses a general dissimilarity measure [51,52]. Within each cluster, we visualized the probability of opening the app during the day over time with 95% CI. The appropriate number of clusters was explored through the *elbow* method and *silhouette* method [53]. The *elbow* method explains the variance of the data in relation to the number of clusters and shows by how much the addition of another cluster would reduce the dissimilarity measure. The *silhouette* method shows how well each user fits into their respective cluster through 2 distance measures: separation (ie, the average distance to the closest other cluster) and compactness (ie, the average within-cluster distance) [54,55]. Kaplan-Meier plots show the estimated cumulative proportion of users engaged and the time scale is days after download [56,57].

Textbox 1. Data visualization methods, data, and engagement measures.

Set A:

- Heat maps: Total count of sessions and total amount of time spent on *Drink Less*, by hour and day of the week
- Timeline plots: Frequency and median amount of time per session
- Kaplan-Meier plots: Time to disengagement (defined as days after download followed by 7 or more consecutive days of nonuse)

Set B:

- K-modes clustering: Was the app used or not each day, over 30 days after download

To explore the association between the delivery of the notification and subsequent near-term engagement of opening the app (ie, engagement in the hour after the notification is delivered), we compared opening the app (yes or no) between the *exposed* time period (11 AM to noon) and an *unexposed* time period (10 AM to 11 AM). We estimated the association between exposure to the notification and opening of the app, which was quantified using a risk ratio. We fitted a marginal model for the outcome of opening the app by using a generalized estimating equation [58] with robust standard errors and an independent working correlation matrix. We fitted an unadjusted model and a model adjusted for the baseline covariates of the continuous variables age, days after download and baseline AUDIT score, which were all included as linear terms, and the categorical variables employment type and gender. Further models explored effect moderation by adding an interaction between exposure to the notification and (1) days after download and (2) cluster (as identified by the k-modes analysis). In the final model, we additionally allowed the association between cluster and exposure to the notification to vary linearly by day after download. Estimated risk ratios with 95% CIs and Wald test *P* values are presented. For models with interaction terms, we present risk ratios for exposure to the notification estimated at days 1, 7, and 30 after download, estimated separately for each cluster.

Results

Overview

The user characteristics are reported in Table 1. Approximately half (49.5%) of the sample were male. The mean age of users was 44 (SD 11.2) years, and the majority worked in nonmanual employment (71.7%). Just under half (46.6%) had a baseline AUDIT score indicating hazardous alcohol consumption (8 to 15, inclusive).

Table 1. User characteristics (N=19,233).

User characteristics	Participants
Sex, n (%)	
Male	9540 (49.60)
Age (years), mean (SD)	44 (11.2)
Employment type, n (%)	
Nonmanual employment	13,792 (71.71)
AUDIT[a] risk zone, n (%)	
Hazardous (8-15)	8958 (46.58)
Harmful (16-19)	3949 (20.53)
At risk of alcohol dependence (20-40)	6326 (32.89)

[a]AUDIT: Alcohol Use Disorders Identification Test.

Summative tables of use (screen views and time on app) by module are provided in Multimedia Appendix 2. It was observed that 85% of screen views occurred in the module *Self-Monitoring and Feedback*. The number of users who reported at least one alcohol-free day or at least one alcohol drink record was 61.86% (11,898/19,233) and 49.11% (9445/19,233), respectively. Over the first 30 days of use after download (derived for Set B data), the median number of sessions per user was 9, with an IQR of 2 to 28 sessions, and the median time spent per user was 24 min, with an IQR of 9 to 55 min.

Visualizations

Patterns of Frequency of Use, Length of Use, Entries of Alcohol-Free Days, and Alcohol Units Consumed

In Figure 1, both heat maps show days of the week along the x-axis and hour of the day along the y-axis. Plot A in Figure 1 shows the frequency of opening the app by hour of the day and day of the week. This shows that there is a strong association between delivery of the notification and opening of the app in the following hour, and this is consistent throughout the week. Plot B in Figure 1 shows the amount of use by hour of the day and day of the week. This shows that the notification is also

associated with the distribution of the total time spent on the app. In plot B, hotspots are observed across the evenings and on Saturday, Sunday, and Monday mornings, which are not evident in plot A. A heat map of when *Drink Less* was downloaded (Multimedia Appendix 3) shows hotspots of downloads on Sunday and Monday evenings. Rotating 3D heat map films of Figure 1, which show the variations more clearly, are provided in Multimedia Appendices 4 and 5.

In Figure 2, plot C shows the median time spent on the app along the y-axis and plot D shows the total number of sessions starting in the hour along the y-axis. Timeline plots show the hour of the day on the x-axis. Plot D shows that the frequency of sessions sharply peaks in the hour after the notification is sent at 11 AM. A second natural peak of frequency occurred in the evenings and a third smaller peak in the mornings. Plot C shows that the median length of time drastically dropped from 11 AM onward, with a slow and steady recovery as the day progressed. An animation of plot D over time is provided in Multimedia Appendix 6, showing that the shape of the distribution over 30 days remains consistent.

Figure 1. Heat maps of total frequency of use (sessions) and total time on app (hours).

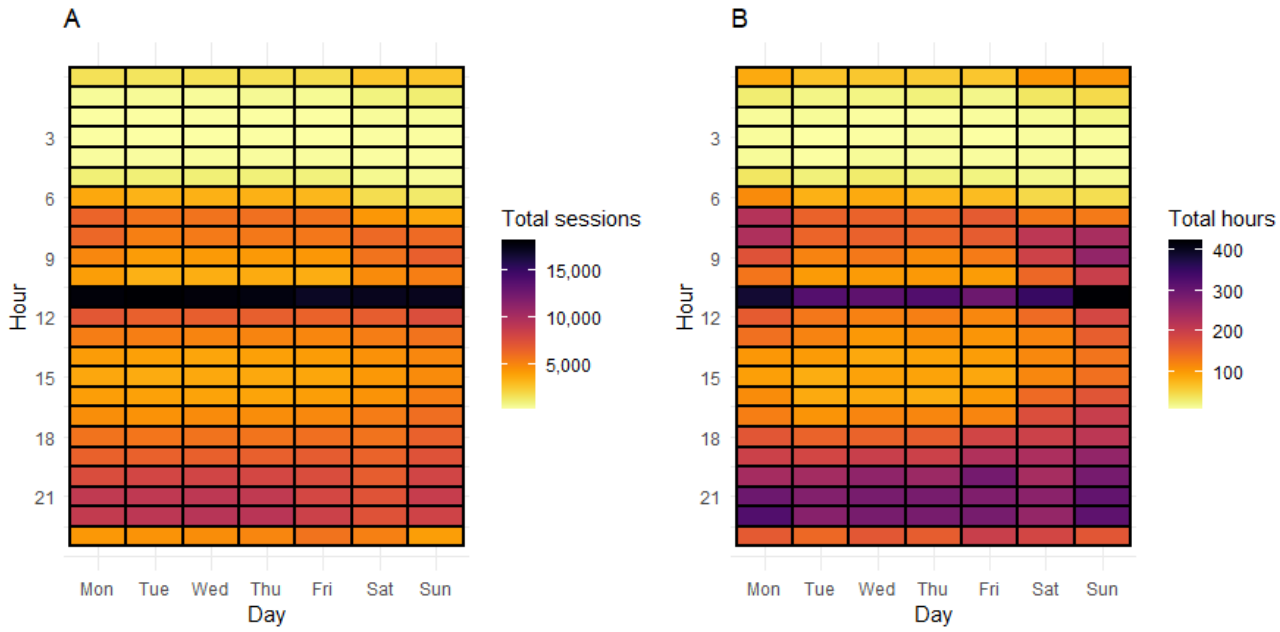
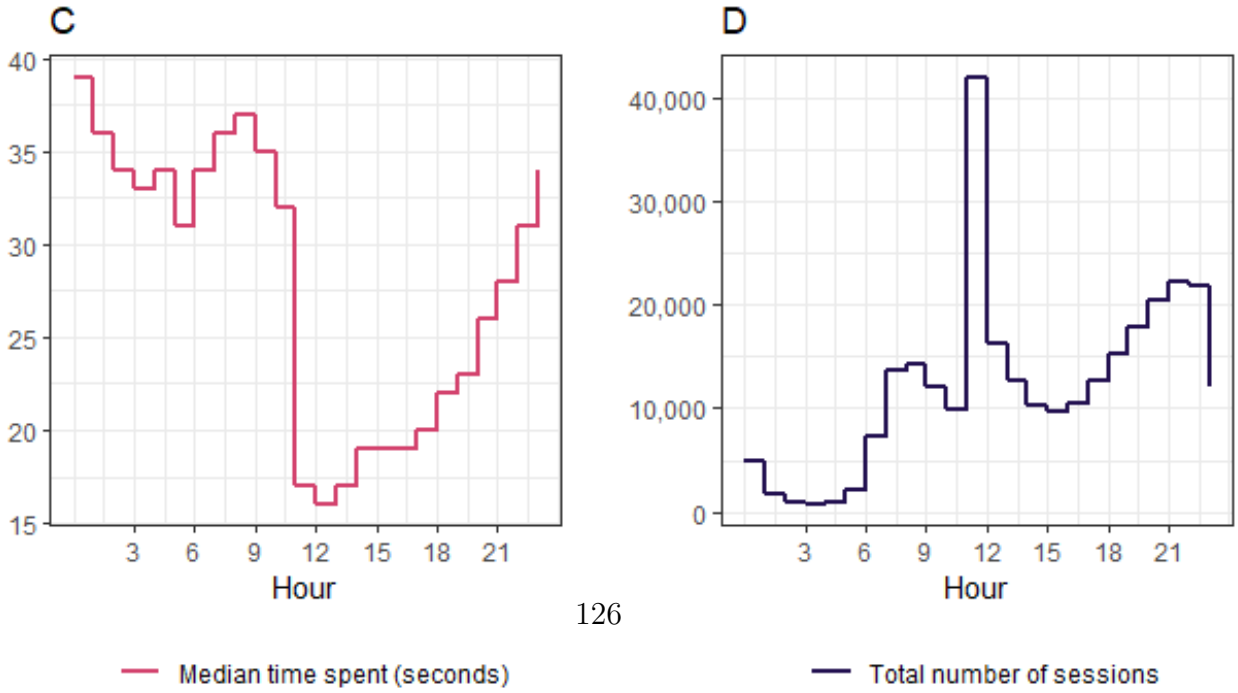
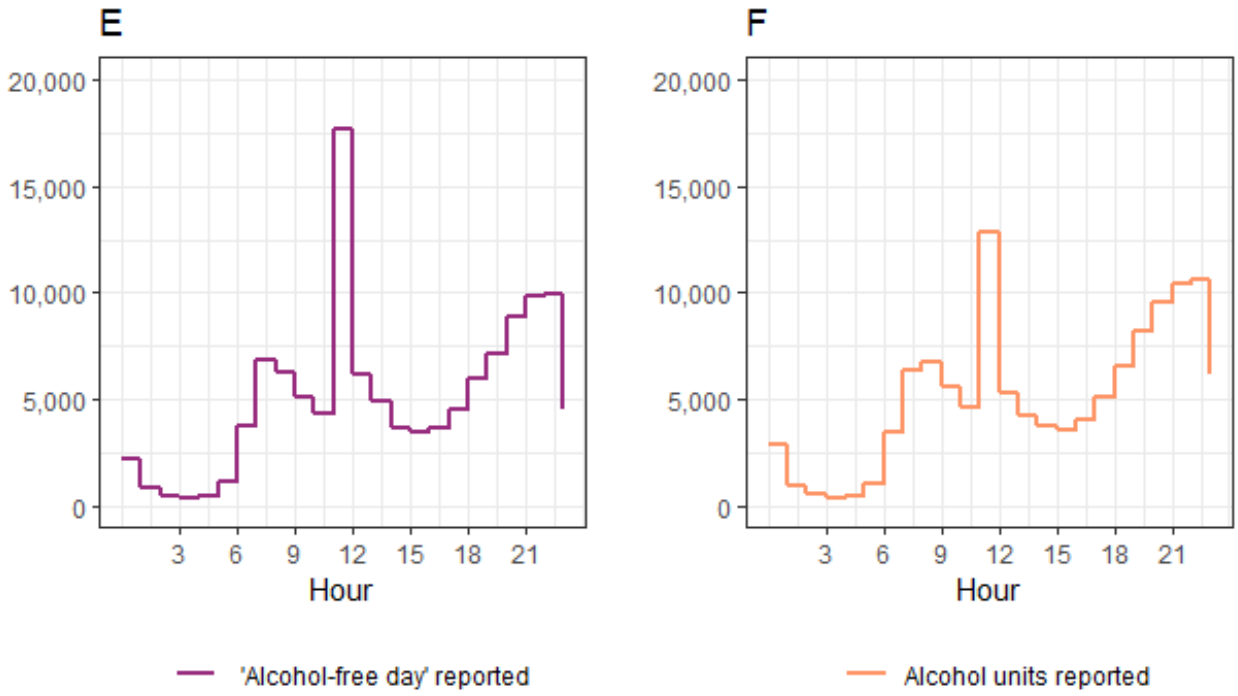


Figure 2. Median time spent on the *Drink Less* app per session and frequency distribution of sessions.



In Figure 3, plot E shows the frequency distribution of entering an alcohol-free day and plot F shows the frequency of entering a drink record. Timeline plots show the hour of the day on the x-axis. There are more alcohol-free days entered between 11 AM to 12 PM than drink records made, which suggests that the notification is more strongly associated with entering *alcohol-free days* than entering *alcohol units consumed*. Both outcomes see similar prominent, natural peaks in the evenings, with an additional smaller peak in the mornings.

Figure 3. Frequency distributions of when alcohol-free days and alcohol units are recorded during the day.



Visualization of Engagement Clusters

A total of 3 clusters emerged from the k-modes clustering. This was based on the measure *did the user open the app?* (binary, yes or no) for the first 30 days after download.

Figure 4 plots the probability of use of the app, stratified by cluster, over time (number of days after download). The 3 ribbons represent the probability of use of the app for each engagement cluster, with 95% CI. On the basis of the observed pattern of engagement, we named the 3 clusters as fast disengagers (10,903/19,233, 56.68%), slow disengagers (3679/19,233, 19.12%), and engagers (4651/19,233, 24.18%). The optimal number of clusters was determined by the elbow method and silhouette method (Multimedia Appendix 7). The silhouette method suggested that the optimal number of clusters was 2, whereas the elbow method suggested 3 clusters. Comparing the results under 2 and 3 clusters showed that the slow disengagers and engagers groups identified under 3 clusters were essentially a subdivision of 1 cluster in the 2-cluster model. We chose to retain 3 clusters based on observed differences in the trajectory of engagement over time between the 2 groups—the engagers and slow disengagers.

The probability of using the app 30 days after download for engagers was 0.69 (95% CI 0.67-0.70), slow disengagers was 0.10 (95% CI 0.10-0.11), and fast disengagers was 0.01 (95% CI 0.01-0.02).

Figure 4. Probability of use on day after download by cluster group with 95% CIs.

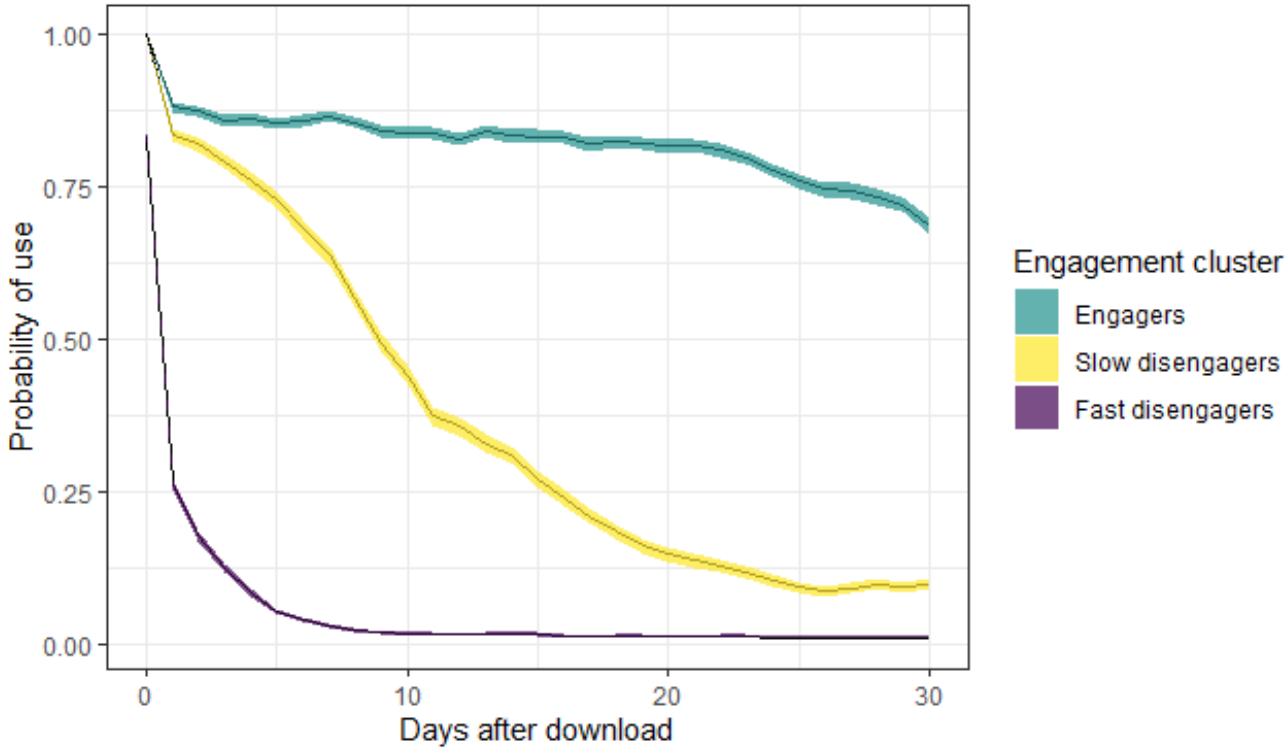


Table 2 shows the distribution of user characteristics across the engagement clusters. The median number of sessions was lowest for the fast disengagers and much higher for the engagers. Engagers are, on average, more likely to be older, male, working in nonmanual employment, and more likely to report a lower AUDIT risk zone, compared with users within the fast disengagers and slow disengagers clusters.

Table 2. User characteristics by cluster group.

User characteristics	Fast disengagers (n=10,903)	Slow disengagers (n=3679)	Engagers (n=4651)
Male, n (%)	4991 (45.78)	1920 (52.49)	2629 (56.53)
Age (years), mean (SD)	43.7 (11.57)	43.2 (10.91)	45.4 (10.57)
Employment type (nonmanual), n (%)	7567 (69.40)	2659 (72.28)	3566 (76.67)
AUDIT[a] risk zone, n (%)			
Hazardous (8-15)	5016 (46.01)	1577 (42.86)	2365 (50.85)
Harmful (16-19)	2155 (19.77)	798 (21.69)	996 (21.41)
At risk of alcohol dependence (20+)	3732 (34.23)	1304 (35.44)	1290 (27.74)
Number of sessions per user, median (25th-75th percentile)	3 (1-6)	18 (12-28)	88 (51-175)

[a]AUDIT: Alcohol Use Disorders Identification Test.

Table 3 provides the estimated associations between exposure to the notification and app use, based on Set B data. Over the first 30 days after day of download, the probability of using the app in the hour after the delivery of the notification was approximately 4 times higher than the probability of using the app in the hour before. All models of the estimated associations between exposure to the notification and app use are adjusted for the continuous variables of age, days after download, baseline AUDIT score, and the categorical variables of employment type and sex. The cluster-specific effects included an effect moderation of the exposure to the notification by cluster group, and the days after download effects included an effect moderation of the exposure by days after download. The adjusted estimated risk ratio was 4.21 (95% CI 4.07-4.36), and the estimated risk ratio was higher among engagers (Wald test *P* value: fast disengagers vs engagers *P*=.001 slow disengagers vs engagers *P*<.001, slow disengagers vs fast disengagers *P*=.44).

Table 4 shows the estimated association between exposure to the notification and opening of the app in the 3 clusters at 3 different time points (days 1, 7, and 30).

Table 3. Estimated associations between exposure to the notification and app use.

Model	Exposure to notification, estimated relative risk ratio (95% CI)
Unadjusted model	4.22 (4.13-4.31)
Adjusted model[a]	4.21 (4.07-4.36)
Days after download[b]	
Day 1	3.93 (3.77-4.10)
Day 7	4.07 (3.93-4.22)
Day 30	4.67 (4.38-4.98)
Cluster[c]	
Fast disengagers	3.97 (3.70-4.25)
Slow disengagers	3.82 (3.60-4.03)
Engagers	4.38 (4.18-4.59)

[a]Adjusted for days after download, employment type, sex, age, and baseline Alcohol Use Disorders Identification Test (AUDIT) score.

[b]Estimated from the model including the interaction effect of exposure to the notification by days after download, adjusted for employment type, sex, age, and baseline AUDIT score.

[c]Estimated from the model including the interaction effect of exposure to the notification by cluster, adjusted for days after download, employment type, sex, age, and baseline AUDIT score.

Table 4. Estimated risk ratio with 95% CI for the associations between exposure to the notification and app use within each cluster, at 3 time points (days 1, 7, and 30).

Clusters	Risk ratio at day 1 (95% CI[a])	Risk ratio at day 7 (95% CI[a])	Risk ratio at day 30 (95% CI[a])
Fast disengagers	3.66 (3.33-4.02)	3.83 (3.57-4.11)	4.58 (3.86-5.43)
Slow disengagers	4.18 (3.85-4.54)	3.87 (3.64-4.12)	2.89 (2.43-3.43)
Engagers	4.05 (3.82-4.30)	4.22 (4.01-4.43)	4.90 (4.56-5.26)

[a]Interaction effect of exposure to the notification and days after download, an interaction effect of exposure to the notification and cluster, and a three-way interaction effect of exposure to the notification, cluster, and days after download, adjusted for employment type, sex, age, and baseline Alcohol Use Disorders Identification Test score.

Visualization of Time to Disengagement

Kaplan-Meier plots, both overall and stratified by clusters, were plotted to show days to disengagement, defined as 7 or more consecutive days of no use, for the first 365 days after downloading *Drink Less*.

In Figures 5 and 6, the x-axis depicts the number of days after download, ranging from 0 to 365, and the y-axis depicts the survival probability, which is the proportion of users who have not disengaged. The dashed lines at the 0.5 survival probability mark shows the time (days) up to when 50%of each cluster has disengaged. Each hash in the plot represents a right-censored user. The number at risk represents the users in the clusters who remain engaged over the year. In Figure 5, we see that 50.00% (9617/19,233) of users have disengaged at 22 days from download, and Figure 6 shows the divergence of longer-term engagement between clusters. The median number of days to disengagement for engagers was 132 days (95% CI 128-137), slow disengagers was 26 days (95% CI 24-29), and fast disengagers was 3 days (95% CI 2-3).

Figure 5. Time to disengagement (defined as the first day of 7 or more consecutive days of no use) for all users.

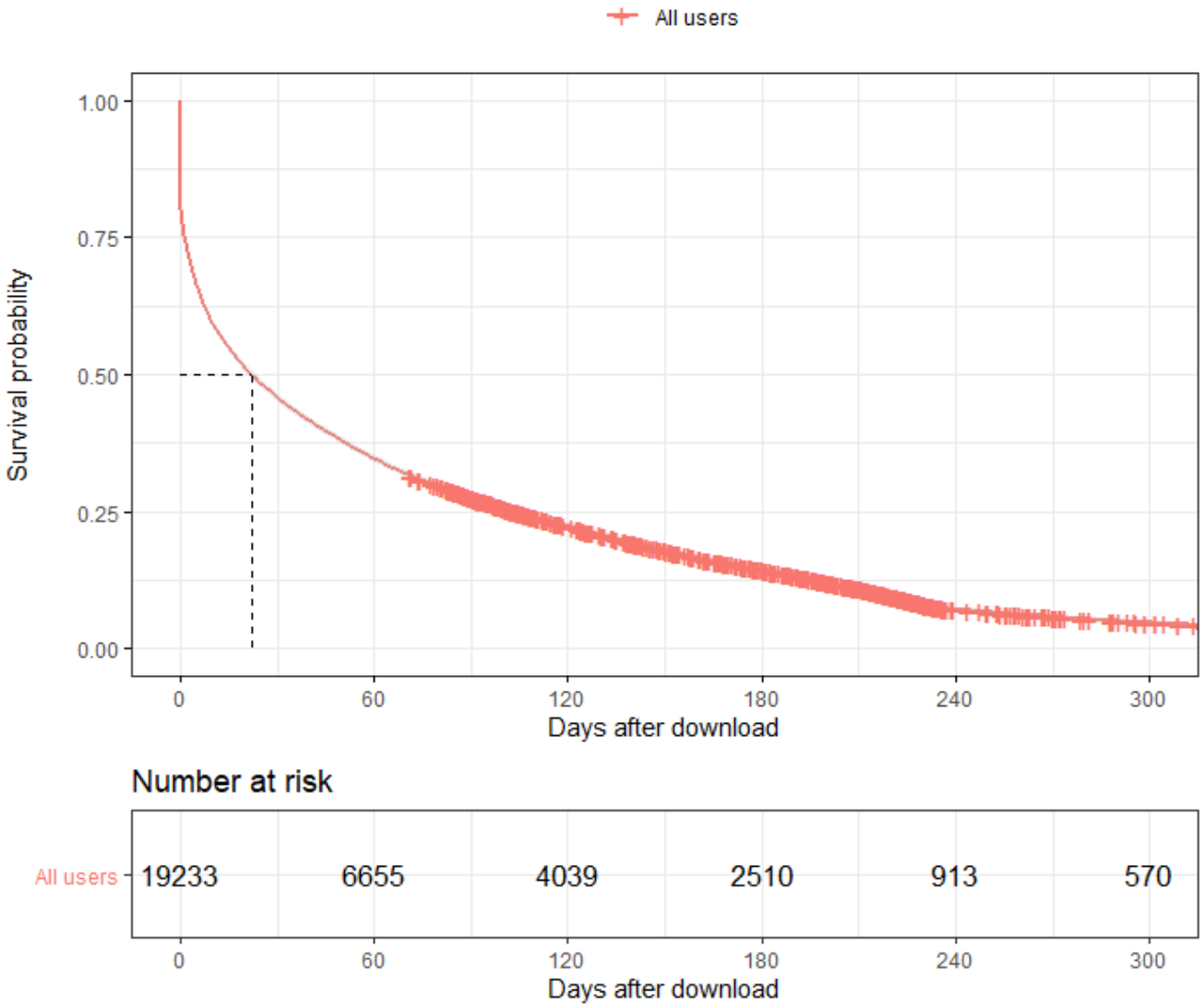
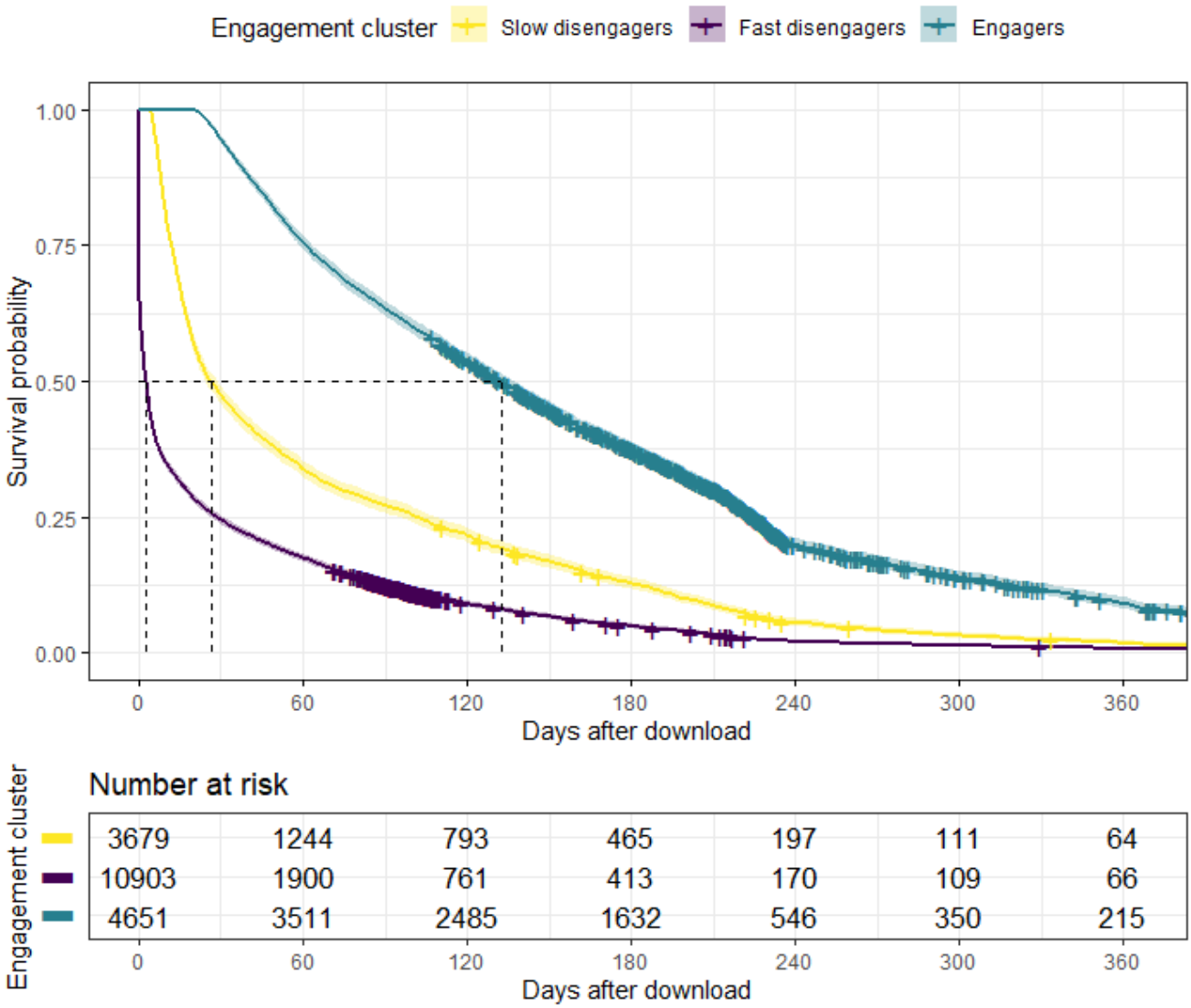


Figure 6. Time to disengagement (defined as the first day of 7 or more consecutive days of no use) by the engagement cluster.



Discussion

Principal Findings

Visualizations provided important insights into how users engage with the behavior change app *Drink Less*. They revealed a strong association between delivery of the daily push notification (sent at 11 AM) and use in the next hour, suggesting that the push notification strongly influences how users engage with *Drink Less* in the immediate hour after the notification is sent. Push notifications (sometimes known as ecological momentary interventions) are programmed messages sent to a user by the app and are commonly employed within behavior change apps to both monitor and provide support to people at risk of harmful alcohol consumption [3,10,59]. Push notifications are a time-varying component of *Drink Less* that can be further optimized to become just-in-time adaptive interventions that rely on decision rules in the provision of real-time support and can learn and adapt to the contextual and psychological circumstances of individuals over time [59]. Previous research has found that notifications are important components that influence engagement with behavior change apps [60-62]. This includes an ecological momentary assessment study with *Drink Less*, which found that establishing a daily routine is important for maintaining engagement and that the daily push notification supports such routines [63]. This study also found that time-varying, endogenous factors of motivation and perceived usefulness of the app were the most consistent predictors of engagement.

Our analysis suggested an approximate adjusted four-fold increase in the probability of using the app in the immediate hour following the notification (11 AM to noon) compared with the preceding hour (10 AM to 11 AM). For the 1 in 5 users belonging to the slow disengagers group, two interesting findings emerged. Firstly, the association between the notification and opening *Drink Less* in the subsequent hour decreased over time, and secondly, patterns of engagement for this group show, on average, a high probability of use over the first week but low probability after the second week. A possible reason for this decline in probability of use and association of the notification and use could be habituation to the daily notification, or turning the notification off. Importantly, we hypothesize that optimizing the notification policy may generate higher rates of engagement for this group.

Future Research to Understand and Optimize the Notification Policy

To carefully create decision rules for the policy to evolve from an ecological momentary intervention to a just-in-time adaptive intervention, we will undertake a micro-randomized trial (MRT). The aim of the MRT is to further develop the push notification policy to improve engagement by targeting internal or external contextual circumstances that either influence excessive drinking (states of vulnerability) or events of engagement with the app (states of acceptability and opportunity) [64]. Visualization of engagement data helped inform the design of our MRT.

Table 5 summarizes how the visualizations from this exploratory research informed the design of our forthcoming MRT.

Primarily, this research guided our decision to shift the delivery time from 11 AM to 8 PM to exploit the potential increase in vulnerability to excess drinking, in an opportune and acceptable moment to engage with *Drink Less* [64]. To avoid the risk of an underpowered MRT, the expected effect size used in the sample size calculation of our MRT is based on a more conservative model, with the control defined as use between 9 AM and 11 AM and the treatment defined as use between 11 AM and 1 PM. This means that the MRT is powered to detect a marginal effect, quantified as a risk ratio, of sending a notification (compared with sending no notification) of 2.16 on user engagement rather than 4.22. We added 2 parallel arms to the MRT to provide an assessment of how engagement with *Drink Less* evolves over time when no notifications are provided and an exchangeable sample to compare the current policy of delivering a fixed notification daily, to a random notification policy, varying the content and sequence of notifications.

Table 5. Linking visualization to the design of a micro-randomized trial.

What we learnt from these analyses	Which visualization or analyses showed us this	How this informed the design of our randomized trial
The present notification appears to be a key driver of engagement	• Figure 1, plot A: heat map of total sessions. • Figure 1, plot B: heat map of total time on app (hours) • Table 4: estimated risk ratio with 95% CI for the associations between exposure to the notification and app use within each cluster, at 3 time points (days 1, 7, and 30)	We chose to undertake a micro-randomized trial to both understand the causal effect of the notification on engagement, and to further optimize the delivery of notifications with respect to time-varying covariates, notifications, and outcomes
The impact of the notification seems to be strongest in the hour preceding delivery	Figure 1, plot A: heat map of total sessions	We set the time window to measure the proximal (ie, near-term) effect as 1 hour after delivery
Evenings seem to be an opportune and acceptable moment to engage with *Drink Less*. It is also a time of increased vulnerability to excess drinking	Figure 1, plot B: heat map of total time on app (hours)	We moved the delivery time of the notification to 8 PM
The notification may encourage the reporting of *alcohol-free days* more than *drink consumed*. This may be due to competing pressures for time at 11 AM	Figure 3: frequency distributions of when alcohol-free days and alcohol units are recorded	We intervened in the evenings to see if this is a more acceptable and opportune time to report drinks consumed
The notification may reduce the median time per session during the reminder of the day	Figure 2: line plot of median time spent on app (seconds)	We included a *no-notification* arm in our trial to capture a momentary assessment of engagement when no notifications are sent
The depth of engagement with *Drink Less* is low	Multimedia Appendix 2: summaries of use by module for all users	We trialed new notifications which target the perceived usefulness of *Drink Less* to encourage broader engagement
Slow disengagers (3679/19,233, 19.13%) have a high probability of engagement during the first week, but by day 30, this group has a low probability, suggesting a loss of motivation	Figure 4: probability of use on day after download by cluster group	We tested 30 new messages to increase novelty and motivation to remain engaged with *Drink Less* (Multimedia Appendix 8)
Exogenous impacts, such as public health campaigns, are likely to influence the cohort of users over time	Figure 6: time to disengagement (defined as the first day of 7 or more consecutive days of no use) by the engagement cluster	We included a *standard app version* arm in the trial, to provide an exchangeable sample to compare the fixed and random notification policies

Limitations

This paper details exploratory research. Our estimates of the association between the notification and opening *Drink Less* do not represent a causal effect on engagement, as we are unable to account for systematic differences in use between the 2 periods that are unrelated to the notification. A randomized trial will allow for the causal effect of the notification to be understood. We also found that simple, accessible visualizations achieved our goal of understanding important patterns of engagement; however, when managing denser streams of data, more complex visualizations may be required.

An additional limitation is that disengagement is defined as a period of no use for 7 or more consecutive days and is considered as a one-time event instead of a repeated event; hence, the Kaplan-Meier plots are interpreted for the survival event *disengagement for the first time*. However, a proportion of users repeatedly disengage and then re-engage with *Drink Less*. It is not uncommon that even after disengaging a number of times, users re-engage for long, continuous spells of use with *Drink Less.* We aim to explore this in future research by visualizing the nature of repeated reengagement with accessible graphical applications and available shared toolsets [34]. An additional limitation is that we did not track whether users subsequently turned off or altered the delivery time of their notifications.

Conclusions

Identifying patterns of engagement from voluminous, temporally dense data presents various challenges for researchers and practitioners. The summarization of such data with heat maps, timeline plots, and Kaplan-Meier plots can provide a clear picture of daily, weekly, and long-term patterns of use over time with a behavior change app. Optimizing engagement is a priority for many behavior change apps, and these visualizations provide a way to identify the key features of how this version of a behavior change app is engaged with.

For *Drink Less*, we have demonstrated the important role of visualizations by showing how these clearly identified how behavioral engagement varies over the day of the week and hour of the day, along with when users first disengage. The visualizations revealed that the daily notification is likely to strongly influence engagement with *Drink Less*. Both the average probability of use over 30 days and the association between use and the notification remained high for users in the engagers cluster yet steadily declined over time for users in the slow disengagers cluster. This suggests that a fixed notification policy can be effective for maintaining engagement for some users but ineffective for others. It is now our priority to understand the causal effect of the notification on engagement and to consider further optimizing the push notification policy to contextual circumstances of individuals over time to inform the development of a just-in-time adaptive intervention. The MRT aims to inform the development of decision rules to tailor the notification policy to individuals over time, with details found in our protocol [65].

Acknowledgments

The development of the *Drink Less* app was funded by the National Institute for Health Research (NIHR) School for Public Health Research, Society for the Study of Addiction, Cancer Research UK, and the UK Centre for Tobacco and Alcohol Studies (UKCTAS). The NIHR School for Public Health Research is a partnership between the Universities of Sheffield, Bristol, Cambridge, and Imperial; University College London; The London School for Hygiene and Tropical Medicine; LiLaC: a collaboration between the Universities of Liverpool and Lancaster; and Fuse—The Centre for Translational Research in Public Health: a collaboration between Newcastle, Durham, Northumbria, Sunderland, and Teesside Universities. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. UKCTAS is part of UK Clinical Research Collaboration, a Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council, and the NIHR under the auspices of the UK Clinical Research Collaboration is gratefully acknowledged.

The authors would like to thank Dr Dave Crane for his important role in the development and factorial screening trial of the *Drink Less* app and the National University of Singapore's Institute of Mathematical Sci*ences* for funding LB and TQ's visit to the program on Statistical Methods for Developing Personalized Mobile Health Interventions. The authors would also like to thank the UCL Tobacco and Alcohol Research Group for their helpful feedback on the draft of this paper.

LB is supported by a PhD studentship funded by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2-R18).

CG is funded by the NIHR, the NIHR School for Public Health Research and Cancer Research UK (CRUK: C1417/A22962).

TQ is supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award number R01AA23187; National Institute on Drug Abuse of the National Institutes of Health under award number P50DA039838; National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54EB020404; and National Cancer Institute of the National Institutes of Health under award number U01CA229437. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

OP is funded by Cancer Research UK (CRUK: C1417/A22962).

EW is supported by the MRC (MR/S01442X/1) and Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), and British Heart Foundation and Wellcome. HP receives funding from Public Health England, the NIHR and NHS England.

Conflicts of Interest

CG is involved in the development and testing of the *Drink Less* app, and OP is an unpaid member of the scientific committee for the Smoke Free app. HP has received consultancy fees from Babylon Health. All other authors declare no conflict of interests.

Multimedia Appendix 1

Privacy notice.
[DOCX File , 14 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Summaries of Use by Module for all users (n=19,233).
[DOCX File , 13 KB-Multimedia Appendix 2]

Multimedia Appendix 3

3D rotating heatmap of when sessions begin.
[MP4 File (MP4 Video), 2999 KB-Multimedia Appendix 3]

Multimedia Appendix 4

3D rotating graph total time on app per session.
[MP4 File (MP4 Video), 3167 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Plot D animated over time.
[MP4 File (MP4 Video), 3436 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Heat map of when downloads occur.
[DOCX File , 29 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Plots of the elbow method and silhouette method.
[DOCX File , 38 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Notification content of new message bank to be trialled in the micro-randomized trial.
[DOCX File , 14 KB-Multimedia Appendix 8]

References

1. UK Chief Medical Officers' Low Risk Drinking Guidelines. Department of Health. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/545937/UK_CMOs__report.pdf [accessed 2020-06-09]
2. Waters AJ, Schoenmakers TM, Snelleman M, Szeto EH, Franken IH, Hendriks VM, et al. Affect, motivation, temptation, and drinking among alcohol-dependent outpatients trying to maintain abstinence: An Ecological Momentary Assessment study. Drug Alcohol Depend 2020 Jan 1;206:107626. [doi: 10.1016/j.drugalcdep.2019.107626] [Medline: 31786398]
3. Beckjord E, Shiffman S. Background for real-time monitoring and intervention related to alcohol use. Alcohol Res 2014;36(1):9-18 [FREE Full text] [Medline: 26258996]
4. Dvorak RD, Pearson MR, Day AM. Ecological momentary assessment of acute alcohol use disorder symptoms: associations with mood, motives, and use on planned drinking days. Exp Clin Psychopharmacol 2014 Aug;22(4):285-297 [FREE Full text] [doi: 10.1037/a0037157] [Medline: 24932896]
5. Cook MA, Newins AR, Dvorak RD, Stevenson BL. What about this time? Within- and between-person associations between drinking motives and alcohol outcomes. Exp Clin Psychopharmacol 2020 Oct;28(5):567-575. [doi: 10.1037/pha0000332] [Medline: 31670547]
6. Chakravorty S, Kuna ST, Zaharakis N, O'Brien CP, Kampman KM, Oslin D. Covariates of craving in actively drinking alcoholics. Am J Addict 2010;19(5):450-457 [FREE Full text] [doi: 10.1111/j.1521-0391.2010.00067.x] [Medline: 20716308]

7. Lac A, Handren L, Crano WD. Conceptualizing and measuring weekend versus weekday alcohol use: item response theory and confirmatory factor analysis. Prev Sci 2016 Oct;17(7):872-881 [FREE Full text] [doi: 10.1007/s11121-016-0685-9] [Medline: 27488456]
8. Lac A, Luk JW. Pathways from positive, negative, and specific alcohol expectancies to weekday and weekend drinking to alcohol problems. Prev Sci 2019 Jul;20(5):800-809 [FREE Full text] [doi: 10.1007/s11121-019-0986-x] [Medline: 30685802]
9. Cohn A, Hunter-Reel D, Hagman BT, Mitchell J. Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011 Dec;35(12):2209-2215 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01571.x] [Medline: 21689119]
10. Morgenstern J, Kuerbis A, Muench F. Ecological momentary assessment and alcohol use disorder treatment. Alcohol Res 2014;36(1):101-109 [FREE Full text] [Medline: 26259004]
11. Petry N, Meredith S, Alessi S. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Technol 2015 Dec:47. [doi: 10.2147/ahct.s65791]
12. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
13. Freisthler B, Lipperman-Kreda S, Bersamin M, Gruenewald PJ. Tracking the when, where, and with whom of alcohol use: integrating ecological momentary assessment and geospatial data to examine risk for alcohol-related problems. Alcohol Res 2014;36(1):29-38 [FREE Full text] [Medline: 26258998]
14. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
15. Forbes G, Newton S, Cantalapiedra Calvete C, Birch J, Dodds J, Steed L, et al. MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial. BMJ Open 2020 Mar 12;10(3):e030164 [FREE Full text] [doi: 10.1136/bmjopen-2019-030164] [Medline: 32165549]
16. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 1;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
17. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
18. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 7;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
19. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mhealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
20. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
21. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. Psychol Methods 2009 Sep;14(3):202-224 [FREE Full text] [doi: 10.1037/a0015826] [Medline: 19719358]
22. Collins LM, Chakraborty B, Murphy SA, Strecher V. Comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. Clin Trials 2009 Feb;6(1):5-15 [FREE Full text] [doi: 10.1177/1740774508100973] [Medline: 19254929]
23. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, et al. The multiphase optimization strategy for engineering effective tobacco use interventions. Ann Behav Med 2011 Apr;41(2):208-226 [FREE Full text] [doi: 10.1007/s12160-010-9253-x] [Medline: 21132416]
24. Anderson R. New MRC guidance on evaluating complex interventions. Br Med J 2008 Oct 22;337(oct22 1):a1937. [doi: 10.1136/bmj.a1937]
25. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
26. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
27. Garnett C, Crane D, West R, Brown J, Michie S. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019 Mar 1;9(2):296-307 [FREE Full text] [doi: 10.1093/tbm/iby043] [Medline: 29733406]

28. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
29. Ola O, Sedig K. The challenge of big data in public health: an opportunity for visual analytics. Online J Public Health Inform 2014;5(3):223 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4933] [Medline: 24678376]
30. Chui KK, Wenger JB, Cohen SA, Naumova EN. Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs. PLoS One 2011 Feb 15;6(2):e14683 [FREE Full text] [doi: 10.1371/journal.pone.0014683] [Medline: 21347221]
31. Ola O, Sedig K. Beyond simple charts: design of visualizations for big health data. Online J Public Health Inform 2016;8(3):e195 [FREE Full text] [doi: 10.5210/ojphi.v8i3.7100] [Medline: 28210416]
32. Gelman A, Pasarica C, Dodhia R. Let's practice what we preach. Am Stat 2002 May;56(2):121-130. [doi: 10.1198/000313002317572790]
33. Ballarini NM, Chiu Y, König F, Posch M, Jaki T. A critical review of graphics for subgroup analyses in clinical trials. Pharm Stat 2020 Sep;19(5):541-560. [doi: 10.1002/pst.2012] [Medline: 32216035]
34. Morrison C, Doherty G. Analyzing engagement in a web-based intervention platform through visualizing log-data. J Med Internet Res 2014 Nov 13;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
35. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A visualization tool to analyse usage of web-based interventions: the example of positive online weight reduction (power). JMIR Hum Factors 2015 May 19;2(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4310] [Medline: 27026372]
36. Sharmin M, Raij A, Epstien D, Nahum-Shani I, Beck JG, Vhaduri S, et al. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:505-516 [FREE Full text] [doi: 10.1145/2750858.2807537] [Medline: 26539566]
37. Chen AT, Wu S, Tomasino KN, Lattie EG, Mohr DC. A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention. J Biomed Inform 2019 Jun;94:103187 [FREE Full text] [doi: 10.1016/j.jbi.2019.103187] [Medline: 31026595]
38. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 1;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
39. Bohn MJ, Babor TF, Kranzler HR. The alcohol use disorders identification test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol 1995 Jul;56(4):423-432. [doi: 10.15288/jsa.1995.56.423] [Medline: 7674678]
40. How a Web Session is Defined in Analytics. Google Analytics. 2017. URL: https://support.google.com/analytics/answer/2731565 [accessed 2020-04-01]
41. Du H, Venkatakrishnan A, Youngblood GM, Ram A, Pirolli P. A group-based mobile application to increase adherence in exercise and nutrition programs: a factorial design feasibility study. JMIR Mhealth Uhealth 2016 Jan 15;4(1):e4 [FREE Full text] [doi: 10.2196/mhealth.4900] [Medline: 26772910]
42. R: A Language and Environment for Statistical Computing. R Core Team. 2020. URL: https://www.R-project.org/ [accessed 2020-11-25]
43. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
44. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, USA: Springer-Verlag; 2016.
45. Morgan-Wall T. Create Maps and Visualize Data in 2D and 3D. CRAN. 2020. URL: https://CRAN.R-project.org/package=rayshader [accessed 2020-04-01]
46. Garnier S. Default Color Maps From 'matplotlib'. CRAN. 2018. URL: https://CRAN.R-project.org/package=viridis, [accessed 2020-04-01]
47. Weihs C, Ligges U, Karsten L, Nils R. klaR analyzing German business cycles. In: Baier D, Decker R, Schmidt-Thieme L, editors. Data Analysis and Decision Support. Berlin, Germany: Springer-Verlag; 2005:335-343.
48. Kassambara A, Kosinski M, Przemyslaw B. Drawing Survival Curves Using 'ggplot2'. CRAN. https://CRAN.R-project.org/package=survminer; 2020. URL: https://CRAN.R-project.org/package=survminer, [accessed 2020-04-01]
49. Lin Pedersen T, Robinson D. A Grammar of Animated Graphics. CRAN. 2020. URL: https://CRAN.R-project.org/package=gganimate [accessed 2020-11-25]
50. Lin Pederson T. The Composer of Plots. CRAN. 2019. URL: https://CRAN.R-project.org/package=patchwork [accessed 2020-04-01]
51. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining Knowl Discovery 1998;2(3):283-304. [doi: 10.1023/A:1009769707641]
52. Bishop CM. In: Jordan M, Kleinberg J, Scholkopf B, editors. Pattern Recognition and Machine Learning. New York, USA: Springer; 2006.
53. Yock AD, Kim G. Technical Note: Using k-means clustering to determine the number and position of isocenters in MLC-based multiple target intracranial radiosurgery. J Appl Clin Med Phys 2017 Sep;18(5):351-357 [FREE Full text] [doi: 10.1002/acm2.12139] [Medline: 28727284]

54. Handl J, Knowles J, Kell DB. Computational cluster validation in post-genomic data analysis. Bioinformatics 2005 Aug 1;21(15):3201-3212. [doi: 10.1093/bioinformatics/bti517] [Medline: 15914541]
55. Lengyel A, Botta-Dukát Z. Silhouette width using generalized mean-a flexible method for assessing clustering efficiency. Ecol Evol 2019 Dec;9(23):13231-13243 [FREE Full text] [doi: 10.1002/ece3.5774] [Medline: 31871641]
56. Morris TP, Jarvis CI, Cragg W, Phillips PP, Choodari-Oskooei B, Sydes MR. Proposals on Kaplan-Meier plots in medical research and a survey of stakeholder views: KMunicate. BMJ Open 2019 Sep 30;9(9):e030215 [FREE Full text] [doi: 10.1136/bmjopen-2019-030215] [Medline: 31575572]
57. Royston P, Parmar MK, Altman DG. Visualizing length of survival in time-to-event studies: a complement to Kaplan-Meier plots. J Natl Cancer Inst 2008 Jan 16;100(2):92-97. [doi: 10.1093/jnci/djm265] [Medline: 18182621]
58. LIANG K, ZEGER SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
59. Bidargaddi N, Schrader G, Klasnja P, Licinio J, Murphy S. Designing m-Health interventions for precision mental health support. Transl Psychiatry 2020 Jul 7;10(1):222 [FREE Full text] [doi: 10.1038/s41398-020-00895-2] [Medline: 32636358]
60. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS ONE 2017 Jan 3;12(1):e0169162. [doi: 10.1371/journal.pone.0169162]
61. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
62. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
63. Perski O, Naughton F, Garnett C, Blandford A, Beard E, West R, et al. Do daily fluctuations in psychological and app-related variables predict engagement with an alcohol reduction app? A series of n-of-1 studies. JMIR Mhealth Uhealth 2019 Oct 2;7(10):e14098 [FREE Full text] [doi: 10.2196/14098] [Medline: 31579022]
64. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
65. Bell L, Garnett C, Qian T, Perski O, Potts HW, Williamson E. Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial. JMIR Res Protoc 2020 Aug 7;9(8):e18690 [FREE Full text] [doi: 10.2196/18690] [Medline: 32763878]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
MRT: micro-randomized trial
NIHR: National Institute for Health Research
UKCTAS: UK Centre for Tobacco and Alcohol Studies

Edited by G Eysenbach; submitted 10.08.20; peer-reviewed by C Morrison, D Leightley, R Krukowski; comments to author 26.08.20; revised version received 08.10.20; accepted 28.10.20; published 11.12.20

Please cite as:
Bell L, Garnett C, Qian T, Perski O, Williamson E, Potts HWW
Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study
J Med Internet Res 2020;22(12):e23369
URL: http://www.jmir.org/2020/12/e23369/
doi: 10.2196/23369
PMID: 33306026

7.1 Supplementary methods

In this supplementary methods section, I describe the details of the K-means [137] and K-modes [101] clustering methodology in this Chapter. This method led me to undercover different trajectories of behavioural engagement with *Drink Less* over the first 30 days since download.

7.1.1 Method description of clustering analysis

The objectives of longitudinal data analysis can be considered as either exploratory or confirmatory [67]. Exploratory analysis generally aims to generate hypotheses by uncovering trends or interesting patterns, commonly through visualisations of the data, whereas confirmatory analysis generally aims to make a statement from the data, by weighing up the evidence in regards to a prespecified established hypothesis.

Statistical learning can be thought of as two broad types of methods: *supervised learning* and *unsupervised learning*. Supervised learning may approach a data set n units with p covariates $X_1, X_2, ..., X_p$ and a response variable Y, with a common goal to predict Y using $X_1, X_2, ..., X_p$ [108]. With unsupervised learning, we do not have the Y predictor variable, and the goal is often to uncover interesting patterns in the data.

Clustering refers to a broad set of methods within unsupervised learning and is primarily used to discover unknown subgroups within the data and gener-

ally performed as part of an exploratory data analysis plan [108]. Clustering can be thought of as method to partition data points such that intra-cluster distances are small compared to the distances between the points that belong to other clusters. The most widely known clustering algorithm is thought to be the unsupervised k-means clustering algorithm [27].

The general goal is to define clusters which minimise the within-cluster variance, and this helps identify heterogeneous sub-groups. K-means clustering is a non-probabilistic, hard-assignment partition-based algorithm which works well with high volume data in a multidimensional space [26]. The terms non-probabilistic technique and hard assignment mean that every data point is assigned to one and only one cluster in each iterative assignment.

The process is initialised with randomly selecting centriods (centers), cycling through a two successive steps of optimisation, iteratively updating the process until convergence. The two stages correspond to the expectation and maximization of the EM algorithm [26], which uses two steps, the E step and the M step [144].

K-means clustering belongs under the family of mixture models and gives a hard assignment to each data point. The process is a type of non-probablistic limit applied to the Expectation-Maximisation algorithm to find the maximum likelihood estimators in latent variable models [26]. Softer methods can be used to introduce a probabilistic approach, which reflects uncertainty over the various assignments of data points to clusters. Such a softer approach

includes fuzzy partitioning, which allows users to belong to more than one group [89].

Clustering falls into a group of probabilistic models known as Gaussian mixture distributions [26]. Superpositions of more than one Gaussian distribution may give a better characterisation of a data set than one distribution alone. Each Gaussian density, called a component of the mixture, has its own mean and covariance in a mixture model.

Below are the steps of how the K-means clustering algorithm proceeds [101, 26].

7.1.2 Notation

Let the goal be to cluster a data set of $\boldsymbol{x_1, ..., x_N}$ points, consisting of $\boldsymbol{N}$ observations of a $\boldsymbol{D}$-dimensional Euclidean variable $\boldsymbol{x}$. The analyst aims to partition the data set into a number of $\boldsymbol{K}$ clusters, where $\boldsymbol{K}$ is pre-specified parameter determined by the analyst.

The $\boldsymbol{K}$-means algorithm is an iterative process which assigns each data point $\boldsymbol{x_n}$ to a cluster $\boldsymbol{K}$. For each data point $\boldsymbol{x_n}$, we set a binary indicator variable $\boldsymbol{r_{nk} \in \{0, 1\}}$ where $\boldsymbol{k = 1, ..., K}$. This tells us which cluster the data point $\boldsymbol{x_n}$ is assigned to. Each $\boldsymbol{K}$ cluster has a summary statistic $\boldsymbol{\mu_k}$ (also known as a *prototype vector* in computer science [26]) calculated from this. We can think of $\boldsymbol{\mu_k}$ as the centre of the associated $\boldsymbol{k^{th}}$ cluster.

The distortion measure $\boldsymbol{J}$ is the objective function, and recall that our overall goal is to minimise $\boldsymbol{J}$ such that the distances between data points in the same cluster are smaller than the distances between data points that are not in the same cluster.

This sets up the overall goal to minimise the squares of the distances of each data point $\boldsymbol{x}$, to its closest vector of $\boldsymbol{\mu_k}$.

The distortion measure $\boldsymbol{J}$ is then defined as an objective function to be minimised.

$$\boldsymbol{J} = \sum_{n=1}^{N} \sum_{k=1}^{K} \boldsymbol{r_{nk}} \|\boldsymbol{x_n} - \boldsymbol{\mu_k}\|^2 \tag{7.1}$$

The K means algorithm proceeds as follows [26]:

1. Initial values for $\boldsymbol{\mu_k}$ are chosen.

2. The first Expectation (E step) is to minimise $\boldsymbol{J}$ with respect to $\boldsymbol{r_{nk}}$ with $\boldsymbol{\mu_k}$ fixed. This is essentially assigning the $\boldsymbol{nth}$ data point to the nearest centre. This can be expressed as

$$\boldsymbol{r_{nk}} = \begin{cases} \mathbf{1} & \text{if k= } \boldsymbol{arg\,min_j} \|\boldsymbol{x_n} - \boldsymbol{\mu_j}\|^2 \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{7.2}$$

3. The next step is to optimise $\boldsymbol{\mu_k}$ with the values of $\boldsymbol{r_{nk}}$ fixed. This is the Maximisation (M step). Recall equation (7.1), which shows the

objective function $\boldsymbol{J}$ is a quadratic function of $\boldsymbol{\mu_k}$. The derivative of $\boldsymbol{J}$ with respect to $\boldsymbol{\mu_k}$ which is set to zero, that is

$$\mathbf{2}\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} \boldsymbol{r_{nk}}(\boldsymbol{x_n} - \boldsymbol{\mu_k}) = \mathbf{0} \tag{7.3}$$

which solves as

$$\boldsymbol{\mu_k} = \frac{\sum_n \boldsymbol{r_{nk}}\boldsymbol{x_n}}{\sum_n \boldsymbol{r_{nk}}} \tag{7.4}$$

4. The EM iterative cycle is repeated until convergence, where convergence means re-allocating data points to a cluster $\boldsymbol{k}$ and re-calculating the cluster means until there is no further change. Each iteration reduces $\boldsymbol{J}$ the distortion measure until it is minimised.

7.1.3 The K-modes clustering algorithm

As stated above, the $\boldsymbol{K}$-means clustering algorithm uses the distortion measure $\boldsymbol{J}$, which is the sum of the squared difference between the cluster mean and the data point. When this data is numerical, this is the Euclidean distance measure, and this prohibits the application of $\boldsymbol{K}$-means to categorical data. When the data is categorical [101], estimating the mean to measure the intra-distance between points is inappropriate.

$\boldsymbol{K}$-modes is an extension to $\boldsymbol{K}$-means [101, 102], which uses both the E and M steps. Three modifications to the $\boldsymbol{K}$-means clustering algorithm resolve the problem that $\boldsymbol{K}$-means can not cluster categorical data-points. The three

modifications are (i) use a matching dissimilarity measure for categorical objects, (ii) replace the centriods summary statistic of the mean with the mode, and (iii) use a frequency-based method to find the modes [101, 102]. This means that the distortion measure $\boldsymbol{J}$ is optimised to reduce the number of disagreements between the data points in the cluster and the mode.

7.2 Conclusion

The use of unsupervised statistical learning with the $\boldsymbol{K}$-modes algorithm, applied to the *Drink Less* cohort data, provided me with a great insight into how the notification policy works for different people. This suggested there was a heterogeneous nature to responding to the notification policy both between users and also within users over time. From this insight, I concluded that optimising the notification policy was a worthwhile investment to improve engagement with *Drink Less*.

In the next section, I detail the protocol of the MRT as well as give further details of the methods of estimation for the primary objective, that is the Estimator for the Marginal Excursion Effect (EMEE).

Chapter 8

Paper Two. Notifications to Improve Engagement with an Alcohol Reduction App: Protocol for a Micro-Randomized Trial

As discussed in the previous paper *Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study*, the exploratory analysis of engagement data with simple and accessible visualisations provided insight to how the fixed notification policy might be

impacting engagement with *Drink Less*. I concluded that understanding the causal effect of the notification on near-term engagement, as well as considering the further optimisation of the notification policy to contextual circumstances over time, was a worthwhile investment.

In this Chapter, I now detail the protocol for the Micro-Randomised Trial (MRT) to achieve these next steps: understanding the near-term causal effect and considering the tailoring to individual contextual circumstances over time.

SCHOOL of HYGIENE &TROPICAL MEDICINE

Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1703883	**Title**	Ms
First Name(s)	Lauren Marie		
Surname/Family Name	Bell		
Thesis Title	Designing Randomised Trials to Improve Engagement through Optimising the Notification Policy of a Behaviour Change App		
Primary Supervisor	Professor Elizabeth Williamson		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	JMIR Research Protocols		
When was the work published?	7th August 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	146
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	LB conceptualised and designed the study with contributions from CG, TQ, OP, HP and EW. CG extracted the data. LB analysed the data. TQ conducted the sample size calculations. LB prepared the manuscript and wrote the first draft. All authors contributed to and approved the final version of the published paper. LB responded to peer review comments with support from all authors.

SECTION E

Student Signature	Lauren Bell
Date	20/06/2022

Supervisor Signature	Elizabeth Williamson
Date	20/06/2022

Protocol

Notifications to Improve Engagement With an Alcohol Reduction App: Protocol for a Micro-Randomized Trial

Lauren Bell[1], MSc; Claire Garnett[2], PhD; Tianchen Qian[3], PhD; Olga Perski[2], PhD; Henry W W Potts[4,5], PhD; Elizabeth Williamson[1,5], PhD

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2]Research Department of Behavioural Science and Health, University College London, London, United Kingdom

[3]Department of Statistics, Harvard University, Cambridge, MA, United States

[4]Institute of Health Informatics, University College London, London, United Kingdom

[5]Health Data Research UK, London, United Kingdom

Corresponding Author:
Lauren Bell, MSc
Department of Medical Statistics
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 02076368636
Email: lauren.bell@lshtm.ac.uk

Abstract

Background: *Drink Less* is a behavior change app that aims to help users in the general adult population reduce hazardous and harmful alcohol consumption. The app includes a daily push notification, delivered at 11 am, asking users to "Please complete your mood and drinking diaries." Previous analysis of *Drink Less* engagement data suggests the current notification strongly influences how users engage with the app in the subsequent hour. To exploit a potential increase of vulnerability of excess drinking and opportunity to engage with the app in the evenings, we changed the delivery time from 11 am to 8 pm. We now aim to further optimise the content and sequence of notifications, testing 30 new evidence-informed notifications targeting the user's perceived usefulness of the app.

Objective: The primary objective is to assess whether sending a notification at 8 pm increases behavioral engagement (opening the app) in the subsequent hour. Secondary objectives include comparing the effect of the new bank of messages with the standard message and effect moderation over time. We also aim to more generally understand the role notifications have on the overall duration, depth, and frequency of engagement with *Drink Less* over the first 30 days after download.

Methods: This is a protocol for a micro-randomized trial with two additional parallel arms. Inclusion criteria are *Drink Less* users who (1) consent to participate in the trial; (2) self-report a baseline Alcohol Use Disorders Identification Test score of 8 or above; (3) reside in the United Kingdom; (4) age ≥18 years and; (5) report interest in drinking less alcohol. In the micro-randomized trial, participants will be randomized daily at 8 pm to receive no notification, a notification with text from the new message bank, or the standard message. The primary outcome is the time-varying, binary outcome of *"Did the user open the app in the hour from 8 pm to 9 pm?"*. The primary analysis will estimate the marginal relative risk for the notifications using an estimator developed for micro-randomized trials with binary outcomes. Participants randomized to the parallel arms will receive no notifications (Secondary Arm A), or the standard notification delivered daily at 11 am (Secondary Arm B) over 30 days, allowing the comparison of overall engagement between different notification delivery strategies.

Results: Approval was granted by the University College of London's Departmental Research Ethics Committee (CEHP/2016/556) on October 11, 2019, and The London School of Hygiene and Tropical Medicine Interventions Research Ethics Committee (17929) on November 27, 2019. Recruitment began on January 28, 2020, and is ongoing.

Conclusions: Understanding how push notifications may impact engagement with a behavior change app can lead to further improvements in engagement, and ultimately help users reduce their alcohol consumption. This understanding may also be generalizable to other apps that target a variety of behavior changes.

International Registered Report Identifier (IRRID): DERR1-10.2196/18690

(JMIR Res Protoc 2020;9(8):e18690) doi: 10.2196/18690

KEYWORDS

mobile health; digital behavior change; engagement; micro-randomized trial; push notifications; excessive alcohol consumption; smartphone app; alcohol; mHealth

Introduction

Excessive alcohol consumption inflicts an array of harms, causing various mental and physical illnesses, loss of productivity, and an increase in violence and traffic accidents [1,2]. There remains a large gap in delivering interventions to at-risk individuals despite the availability of screening and effective interventions [3,4]. As many as four out of five heavy drinkers who attend primary care do not receive screening and brief interventions [5], and this is in part due to barriers to large-scale implementation [6,7]. These barriers include time pressures within general practice, as well as a lack of support and training to successfully shift the consultation from treating the main presenting condition to offering a screening opportunity for hazardous drinking [8].

Behavior change apps promise to reduce this gap by providing real-time data capture and interventions [9-11]. Such behavior change apps, which aim to reduce excessive alcohol consumption, build onto a large body of literature demonstrating the effectiveness of text messages [12-15]. Similar to text messages, behavior change apps can be delivered at a low incremental cost per additional user (high scalability) and offer support to users in real-time. In addition, behavior change apps have the ability to gather data on users, thus enabling them to learn and evolve to become personalized to each individual user. However, a prime challenge for many behavior change apps is poor levels of engagement, with the frequency (number of sessions) and amount (time spent per session) sharply declining over time for the majority of users [16-18].

Engagement with a behavior change app can be considered in two dimensions, behavioral engagement, which can be measured as the amount, frequency, duration, and depth of use, and experiential engagement, characterized by attention, interest, and affect [19]. The behavioral aspect of engagement can be objectively measured through app use data. The effectiveness of a behavior change app can be moderated by a user's engagement with the intervention's active ingredients, and engagement fluctuates within and across users over time [20,21]. A push notification is a message that pops up on the phone, and may also vibrate, make a sound, or lock the screen to gain the user's attention. Push notifications can be sent as a feature to enhance engagement and effectiveness by directing users to engage with the intervention's modules when users likely need it the most [22]. The message can provide a connection between the moments of "point of care" when a user seeks an assessment and intervention, and "point of choice" when a user makes the decision to drink or not to drink [10,23,24].

Drink Less is a behavior change app for the general population of adults seeking to reduce hazardous and harmful alcohol consumption. The research and development of *Drink Less* has been described elsewhere [25-28]. Currently, users receive a daily push notification at 11 am, asking them to "Please complete your mood and drinks diary." Following the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) guidelines [29], the protocol reports our design for a micro-randomized trial (MRT) that aims to improve behavioral engagement (ie, frequency and amount of use) with *Drink Less* through the push notification.

MRT as an Experimental Design for Optimizing Behavior Change Apps

The MRT design is a useful trial design for optimizing the *timing*, *content*, and *sequencing* for push notifications in a behavior change app [30-34]. During an MRT, individuals are repeatedly randomized to actionable notifications, or no notification, at prespecified decision points (Textbox 1). Along with longitudinally measuring a near-term outcome after each decision point, covariate data provided by wearables, sensors, or self-report may also be continuously gathered. Data evolves as a collection of time-varying covariates, treatments, and outcomes. A distinguishing feature of the MRT, compared to a parallel-group randomized controlled trial (RCT), is the repeated randomization over time. This repeated randomization aids further causal inferences that cannot be made when undertaking a parallel-group RCT.

Textbox 1. What does the repeated randomization of notifications in a micro-randomized trial (MRT) offer?

MRT is an experimental design that provides information for developing more optimized policies or decision rules for delivering notifications. The repeated randomization within each individual in an MRT, which is absent in a parallel-group RCT, allows us to understand:

- If the notifications have a near-term effect on engagement, averaging over (i) the course of the study, (ii) all individuals, and (iii) the time-varying contexts individuals experience during the study.
- If the near-term effect of the notifications changes over time or depends on other time-varying covariates of the individual.
- If the notifications have a long-term effect on engagement, in addition to the possible near-term effect.

By understanding the above, researchers can build a more effective and less burdensome policy for delivering notifications, in order to improve users' engagement with a behavior change app [35].

The Drink Less App

Drink Less is a stand-alone app to help people reduce hazardous and harmful alcohol consumption [25-28]. The app was developed in line with the Multiphase Optimisation Strategy Framework [36] and the United Kingdom (UK) Medical Research Council guidance on complex interventions [37]. *Drink Less* contains seven different modules based on behavior change theory and evidence. These modules are (1) Normative Feedback, which is personalized feedback on how an individual's drinking behavior compares to the recommended drinking levels; (2) Goal Setting, which allows users to set weekly "drinking reduction" goals, with brief advice on setting achievable goals; (3) Cognitive bias retraining, delivered through a game which targets users' automatic biases by avoiding cues of alcoholic drinks and approaching nonalcoholic drinks; (4) Self-monitoring and Feedback, which users monitor and reflect on their alcohol consumptions, along with their mood, productivity, sleep and progress on goals; and (5) Action Planning, in which users create plans for dealing with difficult drinking situations. As of January 2, 2020, two new modules were added: (1) Behavioral substitution, which promotes substitution of drinking with a neutral behavior; and (2) Information about Antecedents, which provides users with information about social and environmental situations and events, emotions and cognitions that reliably predict drinking.

Drink Less launched in 2016 and is freely available on iTunes. At onboarding, users are asked to report their age, gender, type of employment (nonmanual, manual, or other) and to complete the Alcohol Use Disorders Identification Test (AUDIT) score [38,39]. The AUDIT is a 10-item screening tool for assessing alcohol consumption that helps identify people who would benefit from reducing or ceasing drinking. Users are also asked why they are using the app ("interested in drinking less alcohol" or "just browsing").

In the existing version of the app, a push notification is sent daily at 11 am asking users to "Please complete your mood and drinks diary." Accessing the app through the notification opens up the app and prompts users to complete their mood and drinks diary.

User Engagement With Drink Less

An Ecological Momentary Assessment study with *Drink Less* users found that establishing a daily routine is important for maintaining engagement and that the daily notification supports such routines [20]. This study also found that perceived usefulness of the app (the belief that using the app will help the user to achieve their goal(s) and an indicator of users' reflective motivation to engage) was associated with increased engagement for some users. The push notification may hence be most effective in improving engagement if it (1) supports the establishment of a routine (being sent at a set time), and (2) motivates to use particular intervention modules.

We visually explored patterns of engagement among a sample of 19,233 existing users of *Drink Less*. Further details of these results are available elsewhere (manuscript submitted). Data analysis from this cohort showed four important findings: (1) use over time decreased, with 50% of users disengaging (no use for seven or more consecutive days) after 22 days since download; (2) the existing daily notification, delivered at 11 am, is likely to have the strongest effect of near-term engagement in the subsequent hour; (3) the breadth of engagement is poor, with 85% of sessions occurring within the "Self-monitoring and Feedback" module; and (4), outside the 11 am notification period, a natural maximum of both frequency and length of sessions appeared in the evenings.

If and how to intervene at "peak-risk" moments is a key research priority [9,10,40,41]. Evenings are a time of day when people with a history of harmful alcohol consumption are the most vulnerable to continued, harmful drinking [42]. Additionally, the visual exploration of engagement patterns over time suggests evenings are an acceptable and opportune moment to engage with *Drink Less*. We decided to exploit the potential increase in vulnerability, opportunity, and acceptability of users in the evenings [43] and to test the marginal effect on near-term engagement of a bank of 30 new push notifications (see Multimedia Appendix 1) delivered at 8 pm. The new messages aim to promote the benefit of using specific intervention modules by targeting users' reflective motivation to use the app.

We will undertake an MRT, with a single decision point of 8 pm, to assess the marginal effect of the new notifications on near-term engagement—use of the app in the hour following the notification—compared with both no notification and to a notification using the existing wording "Please complete your mood and drinking diaries." Within the MRT, we aim to balance the objectives of learning how to optimize the push notification strategy, with the need to trial a good quality app that does not annoy users. Generally, in an MRT, the risk of annoying users with too many notifications over time could be mitigated with lower randomization probabilities. However, there are two reasons why we chose a single decision point to randomize notifications (8 pm), and not test multiple decision points within the day. Firstly, a single decision point allows users to establish an important routine with *Drink Less*, and secondly, this avoids asking users, through the design of the trial, to "Please complete your mood and drinking diaries" more than once within the day.

In order to explore how notifications influence overall engagement, the MRT will be complemented by two parallel trial arms; users will receive the standard notification daily at 11 am in one arm and will receive no notification on any day in the other. The two parallel arms provide us with (1) a momentary assessment of how engagement with *Drink Less* evolves over time when no notification is provided and (2) an exchangeable sample to compare the current policy of delivering a fixed notification daily at 11 am, to randomly varying the content and sequence of notifications at 8 pm.

Aims and Objectives

Aim

This study aims to assess the push notification strategy and to improve engagement with *Drink Less* during the first 30 days following download.

Primary Objective

The primary objective of the study is to estimate the marginal effect of a notification (pooling both types of messages, the standard wording and the new bank of messages) on near-term engagement, defined as the use of the app in the hour following the notification decision point (8 pm to 9 pm).

Secondary Objectives

The secondary objectives are as follows:

1. Compare the marginal effect of the new bank of 30 messages to the standard wording of "Please complete your mood and drinking diary" on near-term engagement, defined as the use of the app in the hour following the notification decision point (8 pm to 9 pm).
2. Explore whether the effect of a notification (pooling both types of messages) on near-term engagement decreases over time.
3. Estimate the lagged effect of prior notifications on near-term engagement.
4. Understand how the notification effect is moderated by time-varying covariates (use before 8 pm, use on the previous day, weekend/weekday effect).
5. Investigate if the effect of the notifications depends on baseline characteristics (gender, age, employment type, AUDIT score).
6. Examine overall engagement during the 30 days following download in users receiving no notifications, those who receive the standard notification daily at 11 am, and those who receive a mix of notifications at 8 pm.

Trial Design

This study is an MRT with two additional parallel arms. Multimedia Appendix 2 illustrates the participant flow through the trial. It also shows which outcomes will be obtained from either the MRT or the two additional trial arms.

The MRT will test the effect of both delivering standard message content and a bank of varied message content on near-term engagement, compared to receiving no message.

Sixty percent of eligible users will be randomly assigned to participate in the MRT. The remaining eligible users will be randomized in equal numbers to the two parallel arms of either receiving no notifications (Secondary Arm A) or daily notification of the standard message of *"Please complete your mood and drinking diary"* (Secondary Arm B).

Among users assigned to the MRT, every day at 8 pm (the "decision point"), each user will be randomized to receive one of three options: no notification, the standard message, or a notification selected at random from the bank of new messages. The randomization probabilities for the decision points each day are 40% to receive no notification, 30% to receive the standard message, and 30% to receive a randomly selected message from the bank of new messages.

Methods

Participants, Interventions, and Outcomes

Study Setting

Drink Less is freely available on the iTunes Store. This trial will recruit eligible new individuals who download the *Drink Less* app during the trial recruitment period, from January 2, 2020, to April 1, 2020 (app version: 2.0.1).

We extended the informed consent process for this trial to comply with ethics requirements. At onboarding, users will be first asked to read the privacy notice and participant information sheet, then provide informed consent (see Multimedia Appendices 2 and 3), before proceeding. Users who do not consent to take part in the research will be provided the standard version of the app.

Eligibility Criteria

Users who download *Drink Less* during the recruitment period will be eligible to participate if they: self-report a baseline Alcohol Use Disorders Identification Test (AUDIT) score of 8 or above, indicating excessive alcohol consumption [39,44]; reside in the UK; are aged 18 years or over, and report themselves to be interested in drinking less alcohol.

Intervention

A bank of 30 novel messages was developed with the aim of increasing users' reflective motivation to engage with the app (see Multimedia Appendix 1). All messages contain the phrase "(using a particular module in the app) *can help you drink less*." As perceived usefulness of the app has previously been found to be associated with increased engagement, we hypothesized that new messages which highlight the benefits of using the app would increase users' reflective motivation to use the app and hence generate higher rates of engagement compared with the standard, existing message.

Measures

Outcome measures for the MRT will be collected continuously over the 30 days following the download of *Drink Less*. These measures collected over time are: when users open the app, when each module is used, the length of time (seconds) spent on the app and drinking records, with the date of drinks consumed and date and time of records made. Outcomes for the wider comparison between the parallel trial arms are defined over the whole 30-day follow-up period.

Outcomes

Primary Outcome (MRT)

The primary outcome measure in the MRT is a time-varying, binary proximal (ie, near-term) measure of engagement (use of the app). Specifically, the primary outcome for the MRT is whether the user opens the app in the hour (8 pm to 9 pm) following the randomization of receiving a notification at 8 pm.

Secondary Outcomes Collected Daily Through the Trial Period (MRT)

Within the MRT, the two secondary outcomes below will be defined in the hour following the decision point (8 pm to 9 pm):

1. whether or not the user creates an entry in their drink calendar (by either recording a drink record or recording a drink-free day);
2. the time, in seconds, spent on the app;

Secondary Outcomes Over the Whole Trial Period (MRT and Parallel Arms)

In the parallel trial arms, secondary outcomes that will be explored are:

1. the number of days to complete disengagement, defined as the first day of at least seven consecutive days of no use from day of download;
2. the total number of sessions over the 30 days following download;
3. the total time, in seconds, spent on the app over 30 days since download, overall, and by intervention module.

Time-Varying Covariates

Measured covariates which vary over time within individuals are the use of the modules (Action Planning; Cognitive Bias Re-Training; Self-monitoring and Feedback; Behavioral Substitution; Goal Setting; Normative Feedback; Information About Antecedents); entry of drink (or alcohol-free) record in each session; if the user opened the app before 8 pm that day; and if the user opened the app the day before.

Time-Fixed Covariates

Measured time-fixed covariates are age, gender, type of employment (manual, nonmanual, or other), day of the week of download, and baseline AUDIT score.

Sample Size

We aim to randomly assign 1200 users to the MRT arm, 400 users to the standard daily notification arm (Secondary Arm A), and 400 users to the no notification arm (Secondary Arm B), resulting in a total of 2000 participants. The sample sizes were calculated as follows:

To estimate the sample size required for the MRT arm, we used a simulation-based approach to determine the sample size required to attain a prespecified power level, because currently there is no off-the-shelf software to calculate the sample size for MRTs with binary outcomes. Our primary objective is to understand the marginal effect of receiving a push notification at 8 pm on engagement, with an important secondary objective towards the tailoring of the notification policy is identifying effect moderation over time. Plausible estimates of a treatment effect and effect moderation were obtained by exploring patterns of use with *Drink Less*. With 80% power and 5% type I error, we have sized this trial to detect a marginal treatment effect of 2.16, which decays by a factor of 0.911 by day since download. This is close to 100% power for our primary objective, the marginal effect. See Multimedia Appendix 4 for more details.

The sample size of the two additional parallel arms was determined based on the secondary outcome of time to disengagement (no use for seven or more consecutive days). Analysis of the current app shows that 55% of users have disengaged by day 30. We powered this sample size based on a minimal relevant change in disengagement of 10%, such that we expect 65% of users to disengage by day 30 when no notifications are delivered. With a 5% type I error and 80% power to detect an increase in disengagement to 65% of users by day 30, we would require 372 users per arm. To simplify the allocation process, we rounded-up the sample size for the parallel arms to 400 users per arm each, resulting in an overall sample size of 2000 and an allocation ratio of 60% to the MRT and 20% to each parallel arm.

Anticipated Recruitment Rate

The available recruitment window with the app was January 2, 2020, to April 1, 2020. All new app users who meet the eligibility criteria, provide consent, and complete app onboarding during this period will be recruited into the trial. Previous analyses of cohort data of existing users suggest that the average number of downloads by eligible users will be 33 per day, through the 59-day recruitment period. If the number of participants exceeds the minimum required sample stated above, we will continue to recruit until the end of the predefined recruitment period.

Assignment of Interventions

Sequence Generation

At recruitment, 60% of participants will be randomized to the MRT. The remaining participants will be randomly allocated 50:50 to receive no notifications or to receive standard notifications daily at 11 am.

Among the participants randomized to the MRT, at 8 pm each participant will be randomized daily to receive one of three options: no notification, the standard notification wording, or a message randomly selected from the new message bank. The randomization probabilities for these three options will be 40%, 30%, and 30%, respectively.

The randomization probabilities are fixed across all individuals and do not depend on individuals' time-varying treatment, outcomes, or covariates.

Allocation Concealment Mechanism

Users will be aware of whether or not they have received a push notification each day. They will be informed in the consent procedures that they are part of a research study testing how different versions of the app affect use. However, they will not receive explicit information that we are interested in the effect of the notification, about which arm of the study they have been allocated to, the full design of the study, or the planned schedule of their notifications. The standard request for users to enable notifications at the end of the onboarding process was disabled for this trial. Users were still able to turn off the notifications through their phone settings.

Implementation

Simple randomization was used, with no stratification or blocking. The code to generate the randomization sequencing was developed and coded into the app by an external app developer. Members of the trial team verified the randomization process.

Data Collection, Management, and Analysis

Descriptions of the trial participants, in terms of their available baseline data, will be reported for all MRT participants and participants in the two additional arms.

Primary Analysis (MRT)

Our primary analysis will estimate the marginal effect of the notifications on the binary, time-varying outcome of whether or not a user opens the app between 8 pm and 9 pm. The marginal effect is averaged over all days and all participants in the MRT arm.

The effect of notifications, quantified as a relative risk, with a 95% confidence interval, will be assessed using the estimator for marginal excursion effect for MRTs with binary outcomes [45]. The excursion effect is a causal effect concerning what would happen if an individual followed the notification scheduled used in the MRT up to day *t*–1 and then deviated from the schedule to receive a notification at day t, versus deviated from the schedule to receive no notification at day *t*. The notification schedule used in the MRT is the delivery of push notifications with 40%, 30%, 30% probability every evening at 8 pm (see the last paragraph of the subsection Trial Design). The marginal excursion effect we consider in the primary analysis will marginalize (ie, average) overall days and all individuals. Because the near-term outcome is binary, we will estimate the marginal excursion effect on the log relative risk scale.

For this analysis, both types of notification—the standard wording and the messages drawn from the new bank of messages—will be pooled; the comparison will be between any notification versus no notification. *P* values less than .05 will be considered statistically significant. Models used in the primary and secondary analyses for the MRT arm will adjust for age, gender, employment type, baseline AUDIT score, the number of days since download, if the user opened the app before 8 pm, and if the user opened the app the day before.

Secondary Analyses (MRT)

Our secondary analyses will assess the effect of sending a notification from the bank of 30 new messages compared to the standard message "Please complete your mood and drinking diaries" on the primary outcome; that is, whether the user opened the app between 8 pm and 9 pm. We will use the same analysis method here as for the primary analyses, which is the estimator for the marginal excursion effect. We will also assess the effect notifications have on users creating an entry to their drinks calendar.

We will investigate the effect moderation of the notification by day in the study, quantified as an interaction, and expressed as a relative risk. We will also examine the sensitivity of the result when day-in-study is replaced by splines or its log-transformation. Lagged notification effects will be similarly quantified.

The continuously valued secondary outcome in the MRT relating to time spent on the app (seconds) will be analyzed using a centered and weighted least-squares estimation method [46] with the effect quantified using the mean difference. All secondary outcomes will be explored by comparing any notification versus none and then separating the two types of notifications.

Secondary Analyses (Parallel Arms)

Time to complete disengagement will be analyzed using the Kaplan-Meier estimator. A Cox proportional hazards model will be used to estimate the hazard ratio for disengagement comparing the three parallel arms. The proportional hazards assumption will be assessed graphically and using tests based on Schoenfeld residuals. If nonproportionality is detected, methods allowing for this will be applied and presented as exploratory analyses alongside the previous Cox model analysis.

Linear regression models with robust standard errors will be used to compare the time spent on the app, both overall and on specific modules, between the three parallel trial arms. Similar models will be used to compare the total number of days of app use between arms.

No adjustment will be made for multiple testing. Outcomes and analyses are categorized by the degree of importance (primary and secondary), and results will be interpreted in the light of that ordering.

Results

This study received funding from the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- R18) in January 2019. As of early March 2020, at the date of manuscript submission, the trial is ongoing, with 452 users recruited.

Data Collection and Data Monitoring

Data collection began on January 2, 2020, and will end on May 1, 2020. Due to the rapid nature of this research, and relatively very low risk of adverse events due to the intervention, there will be no interim analysis or Data Monitoring during the trial.

Ethics and Dissemination

Ethical approval was granted by the London School of Hygiene and Tropical Medicine Interventions Research Ethics Committee (17929) and the University College London Departmental Research Ethics Committee (CEHP/2016/556); an amendment was granted by the Ethics Amendment Request to Work Package One "*The application of digital technologies to advance the understanding, and improve the implementation of behavior change.*"

Confidentiality

No identifiable data will be collected during this study.

Discussion

The study will determine whether sending a notification at 8 pm increases engagement in the subsequent hour with *Drink Less* and whether the impact of the notification changes over time. Previous research has found that the perceived usefulness of the app is a predictor of both the amount and frequency of engagement with *Drink Less* [20]. Building on these findings, secondary analyses will systematically explore if messages which aim to increase the perceived usefulness of the app by encouraging users to try out various modules are more effective at increasing engagement than the standard request to record drinking and mood diary entries. We will also explore potential effect moderation, lagged effects, and overall summaries of use over 30 days since download. This study will provide evidence of how notifications affect engagement, as well as considerations towards further improvement of the push notification policy.

Our research is limited by the lack of outcomes to understand a change in alcohol units consumed, meaning we could not investigate whether receiving notifications had any effect on hazardous and harmful alcohol consumption. Generally, gathering valid and reliable health outcome measures over time, solely through self-reports, is a prime challenge for the digital health community [47,48]. *Drink Less* prompts users to complete the AUDIT-C one month after downloading the app, but the proportion of users to do so is low [49]. Our primary aim is to improve engagement with the app, and future research can investigate whether any effectiveness is mediated through engagement. Importantly, research into effective strategies to collect real-time outcomes on substance abuse through other apps is emerging [50-52], including an MRT with an "engagement-first" strategy to increase the rate of self-reported data [53]. This research is a valuable step towards developing more effective behavior change apps. Another limitation is that we do not understand if users subsequently turned off their notifications during the trial through their phone's settings.

Methodologies for tailoring notification policies, either as a stratified intervention based on time-varying or time-invariant covariates (eg, day of the week, age, past moods, previous app use or drinks reported), or strictly personalized policies, in which user's own responses to prior notifications inform the future policy, are becoming increasingly more refined [54,55]. After establishing whether there is a marginal effect of the push notifications and gaining a better understanding of the push notification's role in the dynamic nature of engagement, subsequent studies may address the more ambitious aims of creating a sequence of decision rules. Such decision rules could capitalize on dynamic states of opportunities within users' current environment or adapt to a user's history. This may be achieved by better understanding how the between- and within-person effects of the notification [56] change under varying circumstances, as well as any lagged effects of notifications.

Acknowledgments

LB is supported by a PhD studentship funded by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2-R18).

CG is funded by the National Institute for Health Research (NIHR) School for Public Health Research and Cancer Research UK (CRUK: C1417/A22962).

OP is funded by Cancer Research UK (CRUK: C1417/A22962).

TQ is supported by National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award number R01AA23187, National Institute on Drug Abuse of the National Institutes of Health under award number P50DA039838, National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54EB020404, and National Cancer Institute of the National Institutes of Health under award number U01CA229437. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

HWWP has received consultancy fees from Babylon Health.

EW is supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome.

Development of the *Drink Less* app was funded by the NIHR School for Public Health Research, Society for the Study of Addiction, Cancer Research UK and the UK Centre for Tobacco and Alcohol Studies (UKCTAS). The NIHR School for Public Health Research is a partnership between the Universities of Sheffield; Bristol; Cambridge; Imperial; and University College London; The London School for Hygiene and Tropical Medicine; LiLaC—a collaboration between the Universities of Liverpool and Lancaster; and Fuse—The Centre for Translational Research in Public Health collaboration between Newcastle, Durham, Northumbria, Sunderland and Teesside Universities. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care." UKCTAS is part of UKCRC, a Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council, and the National Institute for Health Research under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

We would like to thank Dr Dave Crane for his important role in the development and factorial screening trial of the *Drink Less* app, and the National University of Singapore's Institute of Mathematical Sci*ences*, for funding LB and TQ's visit to the program on Statistical Methods for Developing Personalised Mobile Health Interventions. We also thank the University College London Tobacco and Alcohol Research Group, Professor Susan Murphy, and Professor Niranjan Bidargaddi for their helpful feedback on the conceptual development of the trial and draft of this paper.

Authors' Contributions

All authors conceptualized the research and designed the trial. LB acquired project funding and wrote the first draft. OP and CG developed the content of the new notification bank. CG performed data extraction and management for past use of *Drink Less*. LB derived estimates to inform the sample size calcuation. TQ performed sample size simulations. All authors reviewed and edited the manuscript, revising it for intellectual content. All authors gave their final approval for publication and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Bank of 30 newly developed messages and their link to the relevant behavior change module.
[DOCX File , 14 KB-Multimedia Appendix 1]

Multimedia Appendix 2

User flow chart for the micro-randomized trial with two additional parallel arms.
[PNG File , 63 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Privacy notice.
[DOCX File , 14 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Information sheet.
[DOCX File , 14 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Sample size calculation.
[DOCX File , 295 KB-Multimedia Appendix 5]

References

1. Larkin C, Griffin E, Corcoran P, McAuliffe C, Perry IJ, Arensman E. Alcohol Involvement in Suicide and Self-Harm. Crisis 2017 Sep;38(6):413-422. [doi: 10.1027/0227-5910/a000488]
2. Mahnke AH, Miranda RC, Homanics GE. Epigenetic mediators and consequences of excessive alcohol consumption. Alcohol 2017 May;60:1-6. [doi: 10.1016/j.alcohol.2017.02.357]
3. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: An individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec 18;15(12):e1002714. [doi: 10.1371/journal.pmed.1002714]
4. McCambridge J, Rollnick S. Should brief interventions in primary care address alcohol problems more strongly? Addiction 2013 Nov 22;109(7):1054-1058. [doi: 10.1111/add.12388]
5. Anderson P, O'Donnell A, Kaner E. Managing Alcohol Use Disorder in Primary Health Care. Curr Psychiatry Rep 2017 Sep 14;19(11). [doi: 10.1007/s11920-017-0837-z]
6. O'Donnell A, Kaner E. Are Brief Alcohol Interventions Adequately Embedded in UK Primary Care? A Qualitative Study Utilising Normalisation Process Theory. IJERPH 2017 Mar 28;14(4):350. [doi: 10.3390/ijerph14040350]
7. Brown J, West R, Angus C, Beard E, Brennan A, Drummond C, et al. Comparison of brief interventions in primary care on smoking and excessive alcohol consumption: a population survey in England. Br J Gen Pract 2015 Dec 30;66(642):e1-e9. [doi: 10.3399/bjgp16x683149]
8. Rapley T, May C, Frances Kaner E. Still a difficult business? Negotiating alcohol-related problems in general practice consultations. Social Science & Medicine 2006 Nov;63(9):2418-2428. [doi: 10.1016/j.socscimed.2006.05.025]

9. Wray TB, Merrill JE, Monti PM. Using Ecological Momentary Assessment (EMA) to Assess Situation-Level Predictors of Alcohol Use and Alcohol-Related Consequences. Alcohol Res 2014;36(1):19-27 [FREE Full text] [Medline: 26258997]
10. Beckjord E, Shiffman S. Background for Real-Time Monitoring and Intervention Related to Alcohol Use. Alcohol Res 2014;36(1):9-18 [FREE Full text] [Medline: 26258996]
11. Morgenstern J, Kuerbis A, Houser J, Muench FJ, Shao S, Treloar H. Within-person associations between daily motivation and self-efficacy and drinking among problem drinkers in treatment. Psychology of Addictive Behaviors 2016 Sep;30(6):630-638. [doi: 10.1037/adb0000204]
12. Petry N, Meredith S, Alessi S. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. AHCT 2015 Dec:47. [doi: 10.2147/ahct.s65791]
13. Berman AH, Gajecki M, Sinadinovic K, Andersson C. Mobile Interventions Targeting Risky Drinking Among University Students: A Review. Curr Addict Rep 2016 Apr 5;3(2):166-174. [doi: 10.1007/s40429-016-0099-6]
14. Suffoletto B, Kristan J, Callaway C, Kim KH, Chung T, Monti PM, et al. A Text Message Alcohol Intervention for Young Adult Emergency Department Patients: A Randomized Clinical Trial. Annals of Emergency Medicine 2014 Dec;64(6):664-672.e4. [doi: 10.1016/j.annemergmed.2014.06.010]
15. Riordan BC, Conner TS, Flett JAM, Scarf D. A Brief Orientation Week Ecological Momentary Intervention to Reduce University Student Alcohol Consumption. J. Stud. Alcohol Drugs 2015 Jul;76(4):525-529. [doi: 10.15288/jsad.2015.76.525]
16. Ng MM, Firth J, Minen M, Torous J. User Engagement in Mental Health Apps: A Review of Measurement, Reporting, and Validity. PS 2019 Jul;70(7):538-544. [doi: 10.1176/appi.ps.201800519]
17. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Mental Health 2018 Jun 05;21(3):116-119. [doi: 10.1136/eb-2018-102891]
18. Forbes G, Newton S, Cantalapiedra Calvete C, Birch J, Dodds J, Steed L, et al. MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial. BMJ Open 2020 Mar 12;10(3):e030164. [doi: 10.1136/bmjopen-2019-030164]
19. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Behav. Med. Pract. Policy Res 2016 Dec 13;7(2):254-267. [doi: 10.1007/s13142-016-0453-1]
20. Perski O, Naughton F, Garnett C, Blandford A, Beard E, West R, et al. Do Daily Fluctuations in Psychological and App-Related Variables Predict Engagement With an Alcohol Reduction App? A Series of N-Of-1 Studies. JMIR Mhealth Uhealth 2019 Oct 2;7(10):e14098. [doi: 10.2196/14098]
21. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123. [doi: 10.2196/10123]
22. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6. [doi: 10.2196/jmir.4790]
23. Gustafson DH, Shaw BR, Isham A, Baker T, Boyle MG, Levy M. Explicating an Evidence-Based, Theoretically Informed, Mobile Technology-Based System to Improve Outcomes for People in Recovery for Alcohol Dependence. Substance Use & Misuse 2010 Dec 29;46(1):96-111. [doi: 10.3109/10826084.2011.521413]
24. Heron K. E. and J. M. Smyth, Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(Pt 1):1-39. [doi: 10.1348/135910709x466063]
25. Garnett C, Crane D, Michie S, West R, Brown J. Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial. BMC Public Health 2016 Jul 8;16(1). [doi: 10.1186/s12889-016-3140-8]
26. Garnett C. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019;9(2):a-307. [doi: 10.1093/tbm/iby043]
27. Garnett C, Crane D, West R, Michie S, Brown J, Winstock A. User characteristics of a smartphone app to reduce alcohol consumption. Behav. Med. Pract. Policy Res 2017 Mar 17;7(4):845-853. [doi: 10.1007/s13142-017-0477-1]
28. Garnett C, Perski O, Tombor I, West R, Michie S, Brown J. Predictors of Engagement, Response to Follow Up, and Extent of Alcohol Reduction in Users of a Smartphone App (Drink Less): Secondary Analysis of a Factorial Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11175. [doi: 10.2196/11175]
29. Chan A, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013 Jan 09;346(jan08 15):e7586-e7586. [doi: 10.1136/bmj.e7586]
30. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing Digital Integrated Care via Micro-Randomized Trials. Clin. Pharmacol. Ther 2018 Apr 19;104(1):53-58. [doi: 10.1002/cpt.1079]
31. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychology 2015 Dec;34(Suppl):1220-1228. [doi: 10.1037/hea0000305]
32. Law L, Wason J. Use of an embedded, micro-randomised trial to investigate non-compliance in telehealth trials. Trials 2015 Nov 16;16(S2). [doi: 10.1186/1745-6215-16-s2-p222]

33. Klasnja, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
34. Dempsey W, Liao P, Klasnja P, Nahum-Shani I, Murphy SA. Randomised trials for the Fitbit generation. Significance 2015 Dec 10;12(6):20-23. [doi: 10.1111/j.1740-9713.2015.00863.x]
35. Walton A, Collins L, Klasnja P, Nahum-Shani I, Rabbi M, Walton M, et al. The Micro-Randomized Trial for Developing Digital Interventions: Experimental Design Considerations. - 2020 (forthcoming) [FREE Full text]
36. Collins LM, Murphy SA, Strecher V. The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART). American Journal of Preventive Medicine 2007 May;32(5):S112-S118. [doi: 10.1016/j.amepre.2007.01.022]
37. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29:a1655. [doi: 10.1136/bmj.a1655]
38. Bohn MJ, Babor TF, Kranzler HR. The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. J. Stud. Alcohol 1995 Jul;56(4):423-432. [doi: 10.15288/jsa.1995.56.423]
39. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
40. Cohn A. M.., Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment. Alcohol Clin Exp Res 2011;35(12):2209-2215. [doi: 10.1111/j.1530-0277.2011.01571.x]
41. Wright C, Dietze PM, Agius PA, Kuntsche E, Livingston M, Black OC, et al. Mobile Phone-Based Ecological Momentary Intervention to Reduce Young Adults' Alcohol Use in the Event: A Three-Armed Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Jul 20;6(7):e149. [doi: 10.2196/mhealth.9324]
42. Day AM, Celio MA, Lisman SA, Spear LP. Gender, history of alcohol use and number of drinks consumed predict craving among drinkers in a field setting. Addictive Behaviors 2014 Jan;39(1):354-357. [doi: 10.1016/j.addbeh.2013.10.014]
43. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychology 2015 Dec;34(Suppl):1209-1219. [doi: 10.1037/hea0000306]
44. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun 24;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
45. Qian T. Estimating Time-Varying Causal Excursion Effect in Mobile Health with Binary Outcomes. - 2020 [FREE Full text]
46. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing Time-Varying Causal Effect Moderation in Mobile Health. Journal of the American Statistical Association 2017 Mar 31;113(523):1112-1121. [doi: 10.1080/01621459.2017.1305274]
47. Torous J, Firth J. Bridging the dichotomy of actual versus aspirational digital health. World Psychiatry 2018 Jan 19;17(1):108-109. [doi: 10.1002/wps.20464]
48. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Mar 13;35(4):354-362. [doi: 10.1038/nbt.3826]
49. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1). [doi: 10.1038/s41598-018-22420-8]
50. Monk RL, Heim D, Qureshi A, Price A. "I Have No Clue What I Drunk Last Night" Using Smartphone Technology to Compare In-Vivo and Retrospective Self-Reports of Alcohol Consumption. PLoS ONE 2015 May 19;10(5):e0126209. [doi: 10.1371/journal.pone.0126209]
51. Poulton A, Pan J, Bruns Jr LR, Sinnott RO, Hester R. A Smartphone App to Assess Alcohol Consumption Behavior: Development, Compliance, and Reactivity. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e11157. [doi: 10.2196/11157]
52. Smith A, de Salas K, Lewis I, Schüz B. Developing smartphone apps for behavioural studies: The AlcoRisk app case study. Journal of Biomedical Informatics 2017 Aug;72:108-119. [doi: 10.1016/j.jbi.2017.07.007]
53. Rabbi M, Philyaw Kotov M, Cunningham R, Bonar EE, Nahum-Shani I, Klasnja P, et al. Toward Increasing Engagement in Substance Use Data Collection: Development of the Substance Abuse Research Assistant App and Protocol for a Microrandomized Trial Using Adolescents and Emerging Adults. JMIR Res Protoc 2018 Jul 18;7(7):e166. [doi: 10.2196/resprot.9850]
54. Hu X. Personalized Policy Learning using Longitudinal Mobile Health Data. Preprint available 2020 [FREE Full text] [doi: 10.1080/01621459.2020.1785476]
55. Greenewald K. Action Centered Contextual Bandits. Adv Neural Inf Process Syst 2017;30:5973-5981. [doi: 10.1017/9781108571401.024]
56. Qian T, Klasnja P, Murphy SA. Linear Mixed Models with Endogenous Covariates: Modeling Sequential Treatment Effects with Application to a Mobile Health Study. Statistical Science (to appear). - 2020 (forthcoming).

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
MRT: micro-randomized trial
NIHR: National Institute for Health Research
RCT: randomized controlled trial
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials

Edited by G Eysenbach; submitted 12.03.20; peer-reviewed by B Chakraborty, P Wallace, D Leightley, J Kimmerle; comments to author 09.04.20; revised version received 13.05.20; accepted 27.05.20; published 07.08.20

Please cite as:
Bell L, Garnett C, Qian T, Perski O, Potts HWW, Williamson E
Notifications to Improve Engagement With an Alcohol Reduction App: Protocol for a Micro-Randomized Trial
JMIR Res Protoc 2020;9(8):e18690
URL: https://www.researchprotocols.org/2020/8/e18690
doi: 10.2196/18690
PMID: 32763878

Chapter 9

The Estimator for the Marginal Excursion Effect

9.1 Introduction

In the previous chapter, I presented the published protocol of the MRT, which stated that the estimand of the primary objective is the Causal Excursion Effect (CEE), to be estimated by the Estimator for the Marginal Excursion Effect (EMEE).

This chapter explains why the Estimator for the Marginal Excursion Effect (EMEE) was chosen for the MRT protocol.

9.1.1 Background

Data which arises from a MRT is longitudinally rich, consisting of time-varying covariates, multiple treatment occasions and time-varying outcomes. Users can be randomised up to hundreds or thousands of times, and over this course of multiple treatment occasions (i.e. receiving notifications over the first 30 days), there will be many hidden and stochastic factors which influence the outcome. In this Chapter I will discuss the different types of factors to be considered when modelling longitudinal data [78, 67].

The ideal estimator for the MRT is one which consistently estimates the causal effect of the notification for the primary objective and for any secondary objectives. This requires the marginal effect estimate to be consistent when the marginal effect is averaged over a time-varying effect moderator, and when the marginal effect estimate includes in the working model an interaction with the time-varying effect moderator.

This work compares the consistency, precision and coverage probabilities of EMEE and GEE and is based on the data generating model in *Qian, T., Yoo, H., Klasnja, P., Almirall, D., & Murphy, S. A. (2021). Estimating time-varying causal excursion effects in mobile health with binary outcomes. Biometrika, 108(3), 507-527.* The R code for this simulation was obtained (with permission to include in this thesis) from Dr Tianchen Qian's github account: https://github.com/tqian/binary-outcome-mrt.

9.2 Causal Research Questions

In this section I will describe the types of causal research questions MRTs are designed to answer. I will begin by detailing the general steps and framework for answering causal questions under the potential outcomes framework. I will then recap the primary and secondary research questions MRTs often seek to answer and the nature of the data gathered in MRTs. I will then describe the challenges of estimation and inference with such data commonly gathered from an MRT.

9.2.1 General Roadmap for approaching Causal Questions

Here I will follow the general roadmap for approaching causal questions, provided by Petersen and van der Laan [173] which advocates for the following steps and framework to answer causal questions under the potential outcomes framework:

- Define the causal parameter, by describing the notation and data
- Set the causal assumptions and give plausible arguments for or against the causal arguments with respect to the domain knowledge
- Identify the causal parameter
- Proceed to estimation and inference

- Through simulations undertake sensitivity analysis (i.e., how would the results change with different trial designs, different working models or different true data generating mechanisms?)

For the main primary outcome, I test if receiving any notification, pooling both types, increases near-term engagement, compared to receiving no notification. For the secondary objectives to explore effect moderation of the near-term marginal effect, I test if contextual states of interest modify the near term effect. I set the contextual states as habituation, defined as *"did the user receive a notification the day before?* yes/no" and already engaged, defined as *"did the user open the app the day before between 8PM-9PM?* yes/no".

9.2.2 Recap of questions MRTs can answer

Micro-Randomised Trials are experimental designs for the optimisation of a time-varying intervention on a time-varying outcome. The time-varying intervention here is the notification policy. In all MRT designs, the randomisation probabilities are known and established in the trial design and protocol.

Typical questions MRTs aim to address are:

- Does the treatment, on average, work?
- Does the context just prior to receiving a treatment modify the near-

term effect of the notification?

9.2.3 Primary research question of the *Drink Less* MRT

The MRT that I conducted as part of this thesis addresses the following question "Does sending a notification at 8pm increase proximal engagement in the hour thereafter?"

As discussed in Paper One, the visualisations of the engagement data over a 24-hour period showed a natural peak of engagement (opening the app and length of time on the app) in the evenings. This suggested that during the evenings, users are more likely to be in a state of opportunity to engage with the app. I also assume that users are generally more vulnerable to excessive drinking, and intervening at this time of day may improve the behaviour change app.

As stated in the Micro-Randomised Trial (MRT) protocol, I wanted to exploit this natural peak of both frequency of sessions and length of sessions, and established a decision point at 8pm. In order to balance the objectives of gathering information to further optimise the notification policy with the need to trial an app which does not overburden users, I set a 'send limit' [165] of only asking users to *"Please complete your mood and drinks diary"* at a maximum of once per day. That is, only one decision point per day was set.

Establishing a 'send limit' considers how the marginal near-term causal effect depends on the distribution of the notifications at hand, that is, the random notification policy implemented in the MRT. For example, if the random notification policy randomised users to receive a notification which prompted engagement, on average, twice or three times a day, then the near-term causal effect estimated from this more frequent randomisation policy would be different to near-term causal effect estimated the random notification policy in the *Drink Less* MRT. It is possible a randomised notification policy which sends the same message more frequently, and hence an increase in burden, could result in a null near-term marginal effect, while a message sent, on average, 3 times a week, may have a positive near-term marginal effect.

An alternative approach to establishing the send limit is to conduct a stratified MRT, which can adjust the randomisation strategy to previous treatments, or to build in an availability indicator [64, 132].

9.2.4 Secondary research questions of the *Drink Less* MRT

Along with understanding if the intervention (i.e. notification), on average, improves a near-term outcome (i.e. opening the app in the hour thereafter), important secondary and exploratory analysis also aims to understand how a recent prespecified context, just prior to randomisation, may modify the effect of a time-varying intervention on a time-varying outcome, as well as

how the cumulative sequence of interventions may create delayed effects.

Articulating how the dynamic intervention aims to work, in such dynamic contexts can present different types of effects for researchers to consider. Such considerations include hypothesising how internal or external states can modify the near-term marginal effect. For example, internal states such as being in a relaxed state, recently responding to a notification or recent episodes of heavy drinking will likely modify the near-term notification effect. As discovered in the data visualisation, external factors of weekday is associated with different engagement patterns, with longer (length of time) sessions occurring on Sunday and Mondays evenings.

Furthermore, it is important to acknowledge that the potential moderator of a user's current state, such as habituation in the *Drink Less* MRT, can be influenced by the sequence of past treatments and prior states. The user may be in a relaxed state, but the sequence of notifications in the past (either helpful or unhelpful) will influence how the current state modifies the near-term causal effect.

Such rich, sequential data presents different types of effects which are important considerations when analysing data from an MRT. In the *Drink Less* MRT, we are interested in how the effect is modified over time, and also how the effect is modified by recent contextual states of 'already engaged' and 'habituation'. I am interested in these states, as the data visualisation of the large cohort study showed treatment effect heterogeneity over time within

different sub-groups, which suggested some users may habituate to the daily fixed notification policy [18].

These aspects of the MRT provide a pathway for researchers to further optimise a random policy into a dynamic policy, where the dynamic policy aims to be more effective and less burdensome than the previous random policy. In an MRT, users are randomised many times, and this longitudinal nature and rich high dimensional data gathered during an MRT, to answer such causal questions, raises some unique statistical and methodological challenges. The next few paragraphs will describe these challenges.

9.2.5 Endogenous and exogenous effects, confounding and feedback

Types of effects to be considered include endogenous effects, that can be conceptualised as stochastic variables which are influenced by internal, past, within-person factors, while an exogenous variable is stochastic, yet independent of any past states of an individual [67].

In the *Drink Less* MRT, the randomisation probabilities are fixed and exogenous, as they are determined by the study design. Another example of an exogenous variable is seasonal trends, or an exogenous shock, such as the start of a pandemic. If the randomisation probabilities were conditioned on previous outcomes or current contextual states of the user, then the randomisation probabilities are no longer exogenous.

As stated earlier in the thesis, confounding is when there is a variable associated with an exposure and an outcome, which creates a non-causal association between the exposure and the outcome. The randomisation of the notifications ensures there is no confounding in the *Drink Less* MRT. Endogeneity is different to confounding. An endogenous variable in time-series analysis means that a covariate $\boldsymbol{X_t}$ influences both the outcome of interest (i.e. $\boldsymbol{Y_{t+1}}$) and is influenced by outcomes measured earlier on (i.e. $\boldsymbol{Y_{t-1}}$).

In the *Drink Less* MRT, as an example, endogenous covariates may include past patterns of engagement or feeling states (i.e. mood, perceived usefulness of the app) along with the cumulative sequence of notifications delivered randomly. Another example of endogeneity is when the previous outcome (i.e. opening the app yesterday) would influence the near-term outcome for a future time and this is also influenced by past outcomes (i.e. opening the app a few days ago). When the covariate, such as $\boldsymbol{X_{t-1}}$, influences the outcome at $\boldsymbol{Y_t}$, and that outcome $\boldsymbol{Y_t}$ also influences the covariate at $\boldsymbol{X_t}$, this phenomena is sometimes referred to as *feedback* [67].

For MRT data, examples of subtle issues likely to arise include (i) the presence of endogeneity or feedback; and (ii) when the mean of the outcome under no notification is stochastic (i.e. varies randomly) and highly complex, and may be considered difficult (or impossible) to model correctly. Because MRTs measure the near-term outcome over many treatment occasions, this near-term outcome will change in ways we are not able to fully capture or

model through our data due to unknown changing internal or external states impacting the natural patterns of engagement over time, or how the past sequence of treatments, covariates and outcomes influences how the near-term outcome changes for a recent state.

I now describe the estimation of the primary and secondary objectives of the causal excursion [180, 181] and begin by setting the relevant assumptions.

9.3 Why use the Estimator for the Marginal Excursion Effect (EMEE)?

9.3.1 Estimands, estimators and estimates

The concepts of estimands, estimators and estimates is described here, before proceeding with the identification of the estimand of the *Drink Less* MRT.

For studies estimating an intervention effect, the *estimand* is the true treatment effect of an intervention [135]. It is the quantity which the trial aims to estimate. Various estimands may be of interest. For our MRT, the estimand is the Causal Excursion Effect [180].

The *estimator* is the formula to estimate the estimand from the trial's data. In our case, this is the Estimator for the Marginal Excursion Effect (EMEE).

The *estimate* is the value obtained by applying the formula (the estimator) to the realised MRT data.

9.3.2 Assumptions

To express the causal excursion effect in terms of the observed data, I make the following four standard assumptions.

Consistency

The link between potential outcomes and observed outcomes rests on the consistency assumption. Potential outcomes are hypothetical and postulated to exist [31, 98], and the observed outcomes are the realised outcome values after the treatment is assigned [31]. The consistency assumption states that the user's potential treatments, covariates and outcomes under the observed history are the treatments, covariates and outcomes we observe from their data.

Positivity

This assumption states that at each decision point, for every user, the probability of receiving a notification is always strictly greater than zero and strictly less than one. This ensures there is no fully deterministic element to receiving a notification [228].

Sequential ignorability

This assumption requires the notification and the potential outcomes to be conditionally independent given the observed data. That is, there are no unobserved confounders, and the randomised notification groups are essentially exchangeable [127].

Stable unit treatment value (SUTVA)

The SUTVA assumption states that the potential outcomes for one user are not influenced by any other user and their received notifications [228, 98].

9.3.3 Are these assumptions met in the *Drink Less* MRT?

In the *Drink Less* MRT, treatment is randomized with the probability fixed at 0.6. This is bounded away from **0** and **1**, and this ensures the positivity assumption is met. Because the randomisation probabilities are exogenous (are determined only by the trial design), the sequential ignorability assumption is also met.

The consistency assumption may be violated if different iPhones provide different versions of the treatment. For example, older iPhones may have a different look of the notification to newer, sleeker iPhones, however I discount such effects on the outcome to be negligible or very small.

The SUTVA assumption may not hold when there is a personal influence between individuals in the MRT. For example, two people in the same household may be using *Drink Less*. One user may receive the prompt with the message *"Did you know that playing the "yes please, no thanks" game can help you drink less?"* and then discuss the game with the another member of the household, who also then engages with the game between 8pm to 9pm.

Here the first user's treatment allocation affects the second users outcome under no treatment, violating the SUTVA assumption. The plausibility of the SUTVA assumption would need further consideration, particularly when the app includes social media modules where users interact with one another within the app. However, at present, *Drink Less* does not include any community features.

How to assess the plausibility of the SUTVA assumption in the *Drink Less* MRT is a complex problem. One may consider how smartphones can sometimes collect information about a user's location. However, this might require an IP address. In line with the General Data Protection Regulation (GDPR) and privacy consent, data that is personal and identifiable, such as the IP address, is not collected through *Drink Less*. Therefore, in the MRT conducted, it is not possible to empirically assess potential violations of the SUTVA assumption.

In this MRT, I have assumed any potential influences between users is negligible, and if they exist likely lessen the overall marginal near-term effect, and bias the effect in a conservative direction. This is due to the clustering effects.

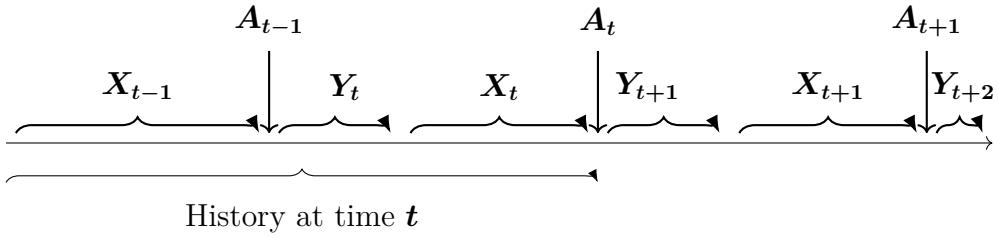
9.3.4 Notation for data in the *Drink Less* MRT

Assume the MRT has $\boldsymbol{n}$ individuals, and there are $\boldsymbol{T}$ decision points in the trial. Both $\boldsymbol{n}$ and $\boldsymbol{T}$ are predetermined by the trial design. The randomised

treatment at time point $\boldsymbol{t}$ is denoted $\boldsymbol{A_t}$. The covariate data collected between decision point $\boldsymbol{t-1}$ and $\boldsymbol{t}$ is denoted $\boldsymbol{X_t}$. The near-term outcome is observed soon after $\boldsymbol{A_t}$ is randomised, and is denoted as $\boldsymbol{Y_{t+1}}$. For each user in the MRT, data evolves as a collection of observed information up to the final decision point at time $\boldsymbol{T}$, with $\boldsymbol{O = (X_1, A_1, Y_2, X_2, A_2, Y_3, ..., X_T, A_T, Y_{T+1})}$.

A timeline and list of details for the notation of the data in the *Drink Less* MRT is provided below.

Figure 9.1: Timeline of MRT data



Time $\boldsymbol{t = 1, 2, \ldots 30}$. There are **30** decision points in the *Drink Less* MRT, one each day over **30** days

Treatment $\boldsymbol{A_t}$ at time $\boldsymbol{t}$: with $\boldsymbol{A_t = 1}$ indicating the delivery of a push notification and $\boldsymbol{A_t = 0}$ indicating no delivery of a push notification.

Covariates $\boldsymbol{X_t}$: Measure of other values, such as Habituation and Already Engaged, as described in the section above.

Outcome $\boldsymbol{Y_{t+1}}$ (Opening *Drink Less* between 8-9pm: binary), measured just after $\boldsymbol{A_t}$ and $\boldsymbol{X_t}$, with $\boldsymbol{Y_{t+1}} = 0$ indicating the app was not opened

and $\boldsymbol{Y_{t+1}} = 1$ indicating the app was opened.

The overbar denotes the sequence of a random variable or realised values up to time $\boldsymbol{t}$, i.e. $\boldsymbol{\bar{A}_t = (A_1, ..., A_t)}$, with the observed history $\boldsymbol{H_t} = \boldsymbol{(\bar{X}_t, \bar{Y}_t, \bar{A}_{t-1})}$ (i.e. everything known at point where $\boldsymbol{A_t}$ is assigned).

States of interest $\boldsymbol{S_t}$, which is a summary of variables from $\boldsymbol{H_t}$, that can be an effect modfier of the near-term causal effect.

Observed Information $\boldsymbol{(\bar{O}_t)}$, up to the final decision point $\boldsymbol{T}$.

9.3.5 Potential Outcomes

I now use the potential outcomes framework to define the effect of the notification on near-term engagement [193]. This framework helps clarify the definition of estimands that describe the causal effect of the notification for both individuals and for populations [67]. I first describe the potential outcomes framework under a simple setting of a point exposure and a single outcome for each individual [98], then extend the potential outcomes framework for estimands which involve notifications expressed as potential outcomes from past notifications [28].

Suppose the intervention is denoted by $\boldsymbol{A}$ and the outcome is denoted by $\boldsymbol{Y}$. The potential outcome notation defines two potential outcomes for each individual, if $\boldsymbol{A}$ is set to $\boldsymbol{a}$, then the potential outcome $\boldsymbol{Y(a)}$. In the MRT example, at the first decision point, a user will have two potential outcomes,

the observed engagement (i.e. opening the app between 8pm to 9pm) if the user received the notification $(\boldsymbol{Y(1)})$ and the value if the user received no notification $(\boldsymbol{Y(0)})$. At any decision point, an individual can not both receive and not receive a notification, hence only one of the outcomes, $(\boldsymbol{Y(1)})$ or $(\boldsymbol{Y(0)})$ is observed. The outcome that would have been observed under the treatment (i.e. notification) that the user did not receive at that precise decision point is the counterfactual (i.e. counter to the fact) outcome.

In our MRT, a notification is randomised each day (i.e there are 30 treatment occasions in my MRT), with a set of potential outcomes, where notifications can be defined as potential outcomes from past notifications [28]. For example, at $\boldsymbol{t = 3}$, the potential outcomes on the second decision point of randomisation has four potential outcomes of $\boldsymbol{Y_3(0,0)}$,$\boldsymbol{Y_3(0,1)}$,$\boldsymbol{Y_3(1,0)}$,$\boldsymbol{Y_3(1,1)}$. Here, $\boldsymbol{Y_3(1,1)}$ is the potential outcome if a user had received a notification on both day 1 and day 2 since download.

9.3.6 Marginal models

In the MRT, we are interested in averages across all variables in the distribution of the randomised policy, that is we are interested in marginal effects.

I will now describe what the term marginal means, and briefly describe well-known two estimators for marginal models: (i) Marginal Structural Models (MSM) and (ii) Sequential Conditional Mean Model (SCMM), then describe why these methods were not suitable to estimate the marginal effect of an

MRT for both primary and secondary objectives.

I will then formally introduce the Estimator for the Marginal Excursion Effect (EMEE), and through simulations, show the estimator's robustness estimate the primary and secondary objectives for an MRT, which I described in the previous sections.

The term marginal is used to emphasise that the model for the mean response on each occasion does not incorporate dependencies on random effects or other user characteristics (such as a user's full history of prior treatments) [78, 67]. To illustrate, marginal, in part, means that the treatment effects do not condition on the individual's full history of data or on the full history of prior treatments. That is, estimates of effect moderation often involve marginalising over all but a small subset of history [180, 181, 78, 67]. The secondary analysis of effect moderation effects in the *Drink Less* MRT is a marginal model conditional on a subset of habituation or 'already engaged' [180]. As such, conditional treatment effects can be estimated which are still based on marginal comparisons [180].

Estimating the parameters in a marginal model requires the specification of three components, (i) the mean response, (ii) the assumptions concerning the variance of the response and (iii) the pairwise within-subject association among the responses [78, 67]. The mean response and the model for the within-subject associations are modelled separately.

Estimators for marginal models can sometimes require that the mean response be correctly modelled [78], and, unlike maximum likelihood estimation [78], do not require any assumptions about the distributions of the observations. However, it is this requirement, that the mean response be correctly modelled for some estimation methods, which can induce challenges for data arising from an MRT, even when the MRT has a fixed randomisation (i.e. when the randomisation probabilities are not influenced by other variables, and there is no confounding) [180].

Sequential conditional mean models fitted using generalised estimating equations

Sequential Conditional Mean Model (SCMM) is a method used for modelling repeated outcomes in longitudinal studies, and can be fitted using a Generalised Estimating Equation (GEE) estimator [114].

Sequential Conditional Mean Model (SCMM) fitted using a Generalised Estimating Equation (GEE) estimator can provide unbiased conditional near-term effects in the presence of endogeneity or feedback under the three following conditions: (i) an independence working correlation matrix is used; (ii) there are no unmeasured confounders; and (iii) the mean model is correctly specified (i.e. for example, if the conditional, marginal model includes a treatment effect moderation term, then this effect moderation term needs to be correctly included.) [181]. If the mean model is correct, then a GEE with

an independence working correlation structure would provide consistent estimates of a conditional effect [181] even if there are endogenous time-varying covariates in the model.

Qian [180] demonstrates that for mean estimates fitted with generalised estimating equations, the estimates will be inconsistent whenever the mean model is misspecified (for example, does not correctly model how the effect is modified over time). This is highly relevant to the *Drink Less* MRT, as we saw through the heatmaps and line plots of engagement over time, we see how engagement patterns are very stochastic and complex. For example, when frequently measuring behavioural engagement with a behaviour change app, there are many stochastic, hidden, unmeasurable, and non-stationary (both exogenous and endogenous) aspects of the user's environment which affect the engagement outcome and can not be fully captured to include in the modelling process.

Marginal Structural Models (MSM) fitted using inverse probability weighted estimation

A common approach for estimating marginal effects from longitudinal studies with time-dependent confounding is Marginal Structural Models (MSM) [114].

MSM is an alternative estimation approach for understanding the causal effect of a time-varying outcome from a time-varying treatment [191]. The

models are generally chosen to analyse longitudinal data when time-varying covariates are considered to be both confounders and intermediate variables in the causal pathways (i.e. feedback is occurring) and are often expressed in terms of counterfactual outcomes. MSM are fitted through the use of inverse probability of treatment weights. The weighting procedure is a two-step process by first creating estimating weights which aim to create a 'pseudo-population', and then secondly allowing estimation to proceed, often without conditioning, to create an unconfounded association between treatments and outcomes.

Estimates of effect moderation can not be accommodated in a standard MSM because the exposure to the treatment and time-dependent covariates are captured in the weights, rather than adjustments which SCMM make. MRTs often seek secondary objectives which examine effect moderation by recent states, and this means the use of MSM are not appropriate for the typical primary and secondary questions asked from an MRT.

I will now describe the estimand the MRT aims to estimate and why a newly proposed estimator can provide consistent estimates for both primary and secondary objectives.

9.4 Causal Excursion Effect (CEE)

The causal excursion effect of $\boldsymbol{A_t}$ on $\boldsymbol{Y_t}$ at time $\boldsymbol{t}$, expressed on the log relative risk scale is:

$$\boldsymbol{\beta_0 = \log \frac{E\{E(Y_t \mid H_t, A_t = 1)\}}{E\{E(Y_t \mid H_t, A_t = 0)\}}} \tag{9.1}$$

This is the estimand targeted in the *Drink Less* MRT.

The Causal Excursion Effect (CEE) is a new type of estimand and the CEE depends on the distribution of the treatments implemented in the MRT. That is, the CEE is a causal effect concerning what would happen if an individual followed the notification scheduled used in the MRT up to day $\boldsymbol{t}$–$\boldsymbol{1}$ and then deviated from the schedule to receive a notification at day $\boldsymbol{t}$, versus deviated from the schedule to receive no notification at day $\boldsymbol{t}$ [180].

The CEE is measured on the log relative risk scale. This is because the risk ratio provides an easy interpretation in terms of exposure to the treatment on the probability of the outcome [90].

The CEE not only provides the causal effect, but can also inform how the current random notification policy implemented in the MRT can be further improved by understanding how varying states modify this marginal causal effect.

In the context of the *Drink Less* example, the primary objective is the marginal effect, which equates to setting $\boldsymbol{S_t = 1}$. The fully marginal excursion effect, $\boldsymbol{\beta_0}$ is defined by

$$\boldsymbol{\beta_0 = \log \frac{E\{E(Y_t \mid H_t, A_t = 1)\}}{E\{E(Y_t \mid H_t, A_t = 0)\}}} \tag{9.2}$$

The effect moderated by $\boldsymbol{S_t}$ is

$$\boldsymbol{\beta_0 + \beta_1 S_t = \log \frac{E\{E(Y_t \mid H_t, A_t = 1) \mid S_t\}}{E\{E(Y_t \mid H_t, A_t = 0) \mid S_t\}}} \tag{9.3}$$

When I examine effect moderation, I set $\boldsymbol{S_t}$ to the realised values of $\boldsymbol{X_t}$ such as habituation or already engaged, as defined above, to understand how the marginal effect is modified by these prior contextual states. Effect moderation analysis can be viewed as contrasts between excursions from the implemented randomised notification policy.

9.4.1 Summary

Up to this point in this chapter, I have characterised the structure of data from an MRT, the assumptions made to infer causal effects, and aspects that estimation methods for marginal effects need to consider, such as endogeneity and feedback. As established in the literature [180, 67, 28, 209], when there is mispecification of the working model, the estimate can be biased if the

generalised estimating equation is used as the estimator. It is also established in the literature that when there is endogeneity, the generalised estimating equation should use an independence working correlation matrix [209, 181].

In this chapter I have formally introduced the Causal Excursion Effect (CEE). The causal excursion effect is a new type of estimand, where the effect depends on the distribution of the treatment at hand, that is the random notification policy implemented in the MRT. The causal excursion effect is the effect of a time-varying treatment which occurs over a defined window of time into the future.

I will now explore the magnitude of the bias found when either the EMEE or GEE is used for estimation, under varying conditions. This chapter now follows with some simulations which further extend the work of Qian [180]. This is in order to understand the performance of EMEE and GEE under various magnitudes of misspecifications with the randomisation probability adopted in the *Drink Less* MRT, and to gain a better understanding of the consistency and efficiency of both the EMEE and GEE(ind) estimators, to model the marginal effect often specified as the primary objective for an MRT. I will examine the performance measures of bias, coverage probability, standard error and root mean standard error for a number of scenarios.

9.5 Simulation Study

This section follows the "ADEMP" structure for simulation work *(Aims, Data-generating mechanism, Estimands, Methods, Performance measures)* [156].

9.5.1 Aim

This work will compare the consistency of two estimators for the marginal treatment effect, EMEE and GEE under model misspecification, exploring the impact of different degrees of effect moderation and different randomisation probabilities.

9.5.2 Data Generating Mechanism

The MRT is a sequential trial with data evolving as a collection of a time-varying treatments $\boldsymbol{A_t}$, covariates $\boldsymbol{Z_t}$ and outcomes $\boldsymbol{Y_t}$.

The total number of decision points ($\boldsymbol{T}$) per individual is 30 (i.e. one decision point over 30 days) for all simulation scenarios. The number of simulated replicate datasets is 1,000 for the sample sizes of 30, 60 and 90.

The time-varying covariate, $\boldsymbol{Z_t}$, is independent of all variables observed before $\boldsymbol{Z_t}$ (i.e. $\boldsymbol{Z_t}$ is not an endogenous covariate and hence no feedback is occurring).

9.5.3 Generation of the treatment

The time-varying treatment $\boldsymbol{A_t}$ is generated from the following Bernoulli distribution, with treatment $\boldsymbol{A_t} = 1$ when a notification is delivered and $\boldsymbol{A_t} = 0$ when a notification is not delivered.

$$\boldsymbol{Pr(A_t = 1) = 0.6} \tag{9.4}$$

9.5.4 Generation of the covariate distribution

The distribution of the time-varying covariate is

$$\boldsymbol{P(Z_t = k) = 1/3,} \text{ for k} \in \boldsymbol{\{0, 1, 2\}} \tag{9.5}$$

9.5.5 Generation of the outcome distribution for each scenario with varying magnitudes of effect moderation

The outcome $\boldsymbol{Y_t}$ is generated from a Bernoulli distribution with

$$\boldsymbol{E(Y_t \mid Z_t, A_t) = \left\{0.2\mathbb{1}_{Z_t=0} + 0.5\mathbb{1}_{Z_t=1} + 0.4\mathbb{1}_{Z_t=2}\right\}e^{A_t(0.1+\gamma Z_t)}} \tag{9.6}$$

$\boldsymbol{Z_t}$ is an effect moderator of the notification effect. $\boldsymbol{Y_t}$ depends on *history* $\boldsymbol{H_t}$ only through $\boldsymbol{Z_t}$ and $\boldsymbol{A_t}$. $\boldsymbol{\gamma}$ is the coefficient which varies the magnitude of effect moderation in each scenario.

Scenario A

The outcome $\boldsymbol{Y_t}$ is generated from a Bernoulli distribution with

$$\boldsymbol{E(Y_t \mid Z_t, A_t) = \left\{0.2\mathbb{1}_{Z_t=0} + 0.5\mathbb{1}_{Z_t=1} + 0.4\mathbb{1}_{Z_t=2}\right\}e^{A_t(0.1+0.3Z_t)}} \quad (9.7)$$

Scenario B

Scenario B eliminates the true effect moderation term by setting $\boldsymbol{\gamma}$ to zero. The outcome $\boldsymbol{Y_t}$ is generated from a Bernoulli distribution with

$$\boldsymbol{E(Y_t \mid Z_t, A_t) = \left\{0.2\mathbb{1}_{Z_t=0} + 0.5\mathbb{1}_{Z_t=1} + 0.4\mathbb{1}_{Z_t=2}\right\}e^{A_t(0.1)}} \quad (9.8)$$

Scenario C

This simulation evaluates the performance measures when $\boldsymbol{\gamma}$ is 0.05.

$$\boldsymbol{E(Y_t \mid H_t, A_t) = \left\{0.2\mathbb{1}_{Z_t=0} + 0.5\mathbb{1}_{Z_t=1} + 0.4\mathbb{1}_{Z_t=2}\right\} e^{A_t(0.1+0.05Z_t)}} \quad (9.9)$$

Scenario D

This simulation evaluates the performance measures when $\boldsymbol{\gamma}$ is 0.10.

$$\boldsymbol{E(Y_t \mid H_t, A_t) = \left\{0.2\mathbb{1}_{Z_t=0} + 0.5\mathbb{1}_{Z_t=1} + 0.4\mathbb{1}_{Z_t=2}\right\} e^{A_t(0.1+0.10Z_t)}} \quad (9.10)$$

9.5.6 Estimands

Estimand for Scenario A

For Scenario A, the true marginal effect (the Estimand) is

$$\boldsymbol{\beta_0 = \log \frac{E\{E(Y_{t,1} \mid H_t, A_t = 1)\}}{E\{E(Y_{t,1} \mid H_t, A_t = 0)\}} = 0.477} \quad (9.11)$$

The true marginal effect (the Estimand) under the data generating mechanism for Scenario A is

$$\boldsymbol{\log \frac{0.2 \times e^{0.1} + 0.5 \times e^{(0.3+0.1)} + 0.4 \times e^{(0.6+0.1)}}{0.2 + 0.5 + 0.4} = 0.477} \quad (9.12)$$

Estimand for Scenario B

The true marginal effect (the Estimand) under the data generating mechanism for Scenario B is

$$\boldsymbol{\beta_0 = \log \frac{E\{E(Y_t \mid H_t, A_t = 1)\}}{E\{E(Y_t \mid H_t, A_t = 0)\}} = 0.10} \tag{9.13}$$

Estimand for Scenario C

The true marginal effect (the Estimand) under the data generating mechanism for Scenario C is

$$\boldsymbol{\beta_0 = \log \frac{E\{E(Y_t \mid H_t, A_t = 1)\}}{E\{E(Y_t \mid H_t, A_t = 0)\}} = 0.16} \tag{9.14}$$

Estimand for Scenario D

The true marginal effect (the Estimand) under the data generating mechanism for Scenario D is

$$\boldsymbol{\beta_0 = \log \frac{E\{E(Y_t \mid H_t, A_t = 1)\}}{E\{E(Y_t \mid H_t, A_t = 0)\}} = 0.221} \tag{9.15}$$

Estimand for Scenario E

The true marginal effect (the Estimand) under the data generating mechanism for Scenario D is

$$\beta_0 = \log \frac{\boldsymbol{E\{E(Y_t \mid H_t, A_t = 1)\}}}{\boldsymbol{E\{E(Y_t \mid H_t, A_t = 0)\}}} = \mathbf{0.221} \tag{9.16}$$

9.5.7 Methods

In this simulation study I assess two methods: the Sequential Conditional Mean Model (SCMM) fitted using the estimator Generalised Estimating Equations which I call GEE for brevity, and the Estimator for the Marginal Excursion Effect (EMEE) (for the marginal effect only).

The definition of a *working model* is a model that is implemented with the awareness that the it may be flawed, such that the working model is not one that will always correctly models the true data generating mechanisms [181, 145]. The working model is the model fitted in the process of obtaining the *estimator*.

Models

For *Scenario A* the working models for the estimator GEE(ind) and EMEE only include the treatment variable. This scenario only models the treatment effect, this is sometimes known as crude model.

Working Models for Scenario B, C, D

For *Scenario B,C and D* the working models for the estimator GEE(ind) and EMEE include the $\boldsymbol{Z_t}$ covariate as a continuous variable. Note the working models are mis-specified as the $\boldsymbol{Z_t}$ covariate is a categorical variables in the data generating mechanism. No interaction effects are fitted in any working models in any of this work.

9.5.8 Performance Measures

The following performance measures will be evaluated:

Bias: The absolute difference between the estimand and the estimate. This is measured as $\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\mathbf{0}}) - \boldsymbol{\beta}_{\mathbf{0}}$ in which $\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\mathbf{0}})$ represents the mean of all estimates across all simulations and $\boldsymbol{\beta}$ represents the true value of the estimand.

Empirical SE: The standard deviation of estimates across all simulated datasets.

Mean Square Error: The expected value of the square of the difference between the estimand and the estimate, calculated as the sum of the variance and squared bias.

Root Mean Square Error (RMSE): Square root of the mean square error.

Coverage Probability: The proportion of replications in the simulations for which the 95% confidence interval of the estimator contained the true value of the estimand. The target values is 95%, i.e. 0.95.

9.5.9 Results

Scenario A

As shown below in Table 9.1, under the crude model (when the working model only includes the treatment effect), no bias is observed in either estimate. The coverage probability, standard deviation and root mean square error are also the same estimates for both estimators.

Table 9.1: Results for Scenario A Bias, Coverage Probability and Precision

Bias & Coverage Probability			
Estimator	Sample Size	Bias	CP
EMEE	30	0.00	0.92
	60	0.00	0.94
	90	0.00	0.95
GEE (ind)	30	0.00	0.92
	60	0.00	0.94
	90	0.00	0.95
Precision			
Estimator	Sample Size	SD	RMSE
EMEE	30	0.08	0.08
	60	0.06	0.06
	90	0.05	0.05
GEE (ind)	30	0.08	0.08
	60	0.06	0.06
	90	0.05	0.05

Scenario B

As shown below in Table 9.2, the results examine the performance measures when the effect moderation term is removed from the true generating mechanism. Neither estimates is biased and the coverage probability, standard deviation and root mean square error are also the same estimates for both estimators.

Table 9.2: Results for Scenario B: Bias, Coverage Probability and Precision

Bias & Coverage Probability			
Estimator	Sample Size	Bias	CP
EMEE	30	0.00	0.94
	60	0.00	0.94
	90	0.00	0.95
GEE (ind)	30	0.00	0.94
	60	0.00	0.94
	90	0.00	0.95
Precision			
Estimator	Sample Size	SD	RMSE
EMEE	30	0.09	0.09
	60	0.06	0.06
	90	0.05	0.05
GEE (ind)	30	0.09	0.09
	60	0.06	0.06
	90	0.05	0.05

Scenario C

As shown below in Table 9.5, the results examine the performance objectives when the true data generating mechanism has the effect moderation $\boldsymbol{\gamma}$ set to 0.05. With this small size in effect moderation, the GEE(ind) and EMEE are consistent, and the coverage probability, standard deviation and root mean square error are similar for both estimators.

Table 9.3: Results for Scenario C: Bias, Coverage Probability and Precision

Bias & Coverage Probability			
Estimator	Sample Size	Bias	CP
EMEE	30	0.00	0.93
	60	0.00	0.95
	90	0.00	0.94
GEE (ind)	30	0.00	0.93
	60	0.00	0.94
	90	0.00	0.94
Precision			
Estimator	Sample Size	SD	RMSE
EMEE	30	0.10	0.10
	60	0.07	0.05
	90	0.06	0.06
GEE (ind)	30	0.08	0.08
	60	0.06	0.06
	90	0.05	0.05

Scenario D

As shown below in Table **??**, the results examine the performance measures when the effect moderation is 0.10. This is an increase to the effect moderation examined in Scenario C (of 0.05). The GEE(ind) is no longer consistent for all sample sizes, and there is potentially some bias starting to appear for the sample sizes of 60 and 90 users.

Table 9.4: Results for Scenario D: Bias, Coverage Probability and Precision

Bias & Coverage Probability			
Estimator	Sample Size	Bias	CP
EMEE	30	0.00	0.94
	60	0.00	0.94
	90	0.00	0.94
GEE (ind)	30	0.00	0.93
	60	0.01	0.94
	90	0.01	0.94
Precision			
Estimator	Sample Size	SD	RMSE
EMEE	30	0.10	0.10
	60	0.06	0.06
	90	0.06	0.05
GEE (ind)	30	0.10	0.10
	60	0.06	0.06
	90	0.05	0.05

Scenario E

As shown below in Table 9.5, the results examine the performance measures when the effect moderation is 0.3 (the same data generating mechanism for Scenario A, with the working model now including the categorical $\boldsymbol{Z_t}$ covariate as a continuous variable). This is an increase of $\boldsymbol{\gamma}$ in the true data generating mechanisms examined in Scenario C and Scenario D. The GEE(ind) estimates are not consistent for all sample sizes. The precision is similar for both estimators.

Table 9.5: Results for Scenario E: Bias, Coverage Probability and Precision

Bias & Coverage Probability			
Estimator	Sample Size	Bias	CP
EMEE	30	0.00	0.94
	60	0.00	0.94
	90	0.00	0.94
GEE (ind)	30	0.03	0.94
	60	0.03	0.92
	90	0.03	0.92
Precision			
Estimator	Sample Size	SD	RMSE
EMEE	30	0.08	0.08
	60	0.06	0.06
	90	0.04	0.04
	30	0.0	0.05
	60	0.05	0.05
	90	0.04	0.05

9.6 Discussion

A common approach to estimating a near-term marginal effect from repeated measures arising in longitudinal data is the GEE, however in this chapter I have described reasons why the GEE with an independence working correlation matrix will not always provided a consistent estimate. The simulations mimic the trial design for the *Drink Less* MRT and provide a strong case for the use of the EMEE for the analysis of the trial.

This chapter provides findings, extending from Qian's work, which suggests that estimates obtained from GEE(ind) are biased when (i) the true data generating mechanism includes a treatment effect moderation (the time-varying outcome is influenced by a time-varying covariate and time-varying treatment interaction) and (ii) the adjusted working model omits or mis-specifies this true relationship, that is, how the time-varying treatment and time-varying covariate interact for the time-varying outcome.

As the data visualisations with patterns of use over time showed, there is likely to be complex effect moderation between dynamic engagement patterns and multiple measures. The estimation approach for the MRT requires a method that guarantees consistency for the marginal treatment effect for binary outcomes, in the presence of such complex and stochastic interactions between a time-varying environmental covariates and time-varying engagement outcomes. The simulations suggest the EMEE guarantees such consistency.

In the simulation studies here, I found that under different alterations to the true data generating mechanism or the working model, the EMEE was always consistent for the marginal effect. The simulations also show that the estimators have similar performance for efficiency measures.

The thesis will now follow with the results of the *Drink Less* MRT which were estimated with the EMEE.

Chapter 10

Paper Three. Optimising the Notification Policy to Improve Engagement with an Alcohol Reduction App: Results from a Micro-Randomized Trial

In the previous chapters I have explored past engagement data with *Drink Less*, which led to the conclusion that optimising the notification policy was a worthwhile investment, and presented a published protocol of the MRT.

In this chapter, I present the results and findings from the MRT for *Drink Less*.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1703883	**Title**	Ms
First Name(s)	Lauren Marie		
Surname/Family Name	Bell		
Thesis Title	Designing Randomised Trials to Improve Engagement through Optimising the Notification Policy of a Behaviour Change App		
Primary Supervisor	Professor Elizabeth Williamson		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Journal of Medicial Internet Research
Please list the paper's authors in the intended authorship order:	Lauren Bell, Claire Garnett, Yihan Bao, Zhaoxi Cheng, Tianchen Qian, Olga Perski, Henry WW Potts and Elizabeth Williamson
Stage of publication	**Undergoing revision**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	LB collected and transformed the data, and analysed the results. TQ provided the code for the Estimator for the Causal Excursion Effect with his students Yihan Bao and Zhaoxi Cheng. LB wrote the first draft and all co-authors contributed to final, submitted manuscript.

SECTION E

Student Signature	Lauren Bell
Date	20/06/2022

Supervisor Signature	Elizabeth Williamson
Date	20/06/2022

How Notifications affect Engagement with a Behaviour Change App: Results from a Micro-Randomised Trial.

Lauren Bell[1], Claire Garnett[2], Yihan Bao[3], Zhaoxi Cheng[3], Tianchen Qian[3], Olga Perski[2], Henry WW Potts[4,5] and Elizabeth Williamson[1,5]

1. Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK.

2. Research Department of Behavioural Science and Health, UCL, London, UK.

3. Department of Statistics, University of California Irvine, USA.

4. Institute of Health Informatics, UCL, London, UK.

5. Health Data Research UK, London, UK.

Abstract

Background *Drink Less* is a behaviour change app to help higher risk drinkers in the UK reduce their alcohol consumption. The app includes a daily notification, asking users to *"Please complete your drinks and mood diary"*, yet we did not understand the causal effect of the notification on engagement nor how to improve this component of *Drink Less*. We developed

a new bank of 30 new messages to increase users' reflective motivation to engage with *Drink Less*. In this study we aimed to determine how both the standard and new notifications affect engagement.

Objective Our objective was to estimate the causal effect of the notification on near-term engagement, to explore whether this effect changed over time, and to create an evidence base to further inform optimisation of the notification policy.

Methods We conducted a Micro-Randomised Trial (MRT) with two additional parallel arms. Inclusion criteria were *Drink Less* users who; (1) consent to participate in the trial; (2) self-report a baseline Alcohol Use Disorders Identification Test score of 8 or above; (3) reside in the United Kingdom; (4) age 18 years and (5) report interest in drinking less alcohol. Our MRT randomised 350 new users to test if receiving a notification, compared to receiving no notification, increased the probability of opening the app in the subsequent hour, over the first 30 days since downloading *Drink Less*. Each day at 8 PM, users were randomised with 30% probability to receive the standard message, 30% probability to receive a new message or 40% probability to receive no message. We additionally explored time-to-disengagement, with the allocation of sixty percent of eligible users randomised to the MRT (n=350), and forty percent of eligible users randomised in equal number to the two parallel arms, either receiving the no notification policy (n=98) or the standard notification policy (n=121). Ancillary analyses explored effect

moderation by recent states of habituation and engagement.

Results Receiving a notification, compared with not, increased the probability of opening the app in the next hour by 3.5-fold (95% confidence interval (CI) 2.91, 4.25). Both message types were similarly effective. The effect of the notification did not change significantly over time. A user being in a state of 'already engaged' lowered the new notification effect by 0.80 (95% CI 0.55, 1.16), though non-significantly. Across the three arms, time-to-disengagement was not significantly different.

Conclusion We found a strong near-term effect of engagement on the notification but no overall difference in time to disengagement between users receiving the standard fixed notification, no notification at all, or the random sequence of notifications within the MRT. The strong near-term effect of the notification presents the opportunity to target notifications to increase 'in-the-moment' engagement. To improve longer-term engagement, further optimisation is required.

International Registered Report Identifier (IRRID): DERR1-10.2196/18690

10.0.1 Introduction

Hazardous and harmful alcohol consumption is one of the major risk factors for many disease outcomes and has a significant global burden of health [8, 81]. Delivering brief interventions to reduce hazardous and harmful alcohol

drinking is known to be effective [110] however such efforts are challenged by the sheer prevalence of harmful drinking and the limited capacity of services [229, 111]. There is a long-standing recognition of the need to broaden the reach of and access to brief, effective interventions to reduce harmful alcohol consumption for help-seeking individuals [36]. A promising solution is behaviour change apps, as these are complex interventions which can capture dynamic patterns in human behaviour and deliver support when an individual needs this the most [82, 14, 146]. Building on evidence which supports short message services as interventions to help individuals [21], behaviour change apps can provide comprehensive, every-day support, within people's homes and diverse communities, to maintain healthy behaviours [194]. However, a major concern is that insufficient engagement with an app is likely to hinder behaviour change, particularly if a user disengages with the app not long after downloading it [234, 4]. Engagement, a construct of both experiential and behavioural aspects [14], fluctuates within and between users over time, and is influenced not only by the static content of the intervention, but also by internal (e.g., the user's momentary mood, cognitive state and recent patterns of engagement and drinking) and external (e.g., the user's current environment) factors [171, 160, 6]. Push notifications (reminders or pop-up messages on the screen) are often implemented to increase engagement with a behaviour change app [4] and can have small, positive effects on engagement over a 24-hour period [25]. However, a more immediate causal effect (e.g., within the next hour) of a push notification on engagement with

behaviour change apps is not yet established [25, 227]. We undertook a trial to estimate the causal effect of the notification on near-term engagement in the behaviour change app *Drink Less* and to consider how the notification policy could be further optimised to improve engagement.

10.0.2 The *Drink Less* App

Drink Less is a behaviour change app that aims to help higher risk drinkers in the UK adult population reduce their alcohol consumption. The app is freely available to people seeking help with their alcohol consumption though the app has not been advertised or targeted to specific groups of people. *Drink Less* was developed in line with the Medical Research Council guidelines for developing and evaluating a complex intervention [40, 204] and the MOST (Multiphase Optimisation Strategy) framework [54, 55], and is freely available on the Apple App Store. *Drink Less* is an evidence- and theory- informed intervention with several modules. The overall development and refinement of *Drink Less*, including how the behaviour change modules were selected, can be found here [88, 85, 84]. The standard version of the app delivers a local daily notification at 11 AM, asking the user to *"Please complete your mood and drinks diary"* (See Appendix 6 for a visual of the *Drink Less* notification). The daily notification aims to remind users to self-monitor their drinking. The National Institute for Health and Care Excellence (NICE) for the United Kingdom recommends self-monitoring as an effective technique for the act of noticing recent behaviour and how this relates to their

related goals [164]. However, if a user has already engaged with the app to self-monitor their drinking that day, the notification may be an unnecessary reminder and ultimately annoy the user over time. The notification appears on the users' Notification Centre and tapping the notification opens to the *Drink Less* landing page. The standard version of *Drink Less* sends a daily notification that aims to increase self-monitoring through tracking of recent alcohol units consumed (i.e., the day before). The 11 AM time is to allow users time to complete their morning routines before engaging with the app. User feedback was received via the App Store, with a suggestion that a reminder to report drinking diaries in the evenings would be more helpful.

10.0.3 Engagement with *Drink Less*

We previously reported exploratory research which visualised the temporal patterns of engagement with *Drink Less* [18]. The visualisations showed limited depth of engagement, with 85% of sessions occurring within the Self-Monitoring and Feedback module, and a natural peak, near 8 PM, of both frequency (i.e., number of logins) and time spent on the app observed in the evenings. This suggested that evenings are opportune moment to engage with *Drink Less* for longer sessions. In the evenings, users may be more vulnerable to harmful drinking and intervening at this moment of vulnerability, and at an opportune moment to engage, may be more conducive to reduce harmful patterns of drinking. Additionally, our exploratory research discovered different trajectories of use, with 50% of users disengaging with the app

22 days after download, and we hypothesised that a fixed notification policy may suit some users for maintaining engagement, while other users may habituate to the daily notification policy and disengage sooner.

10.0.4 Specific aims and objectives

We conducted a Micro-Randomised Trial (MRT), a design in which users recruited to the MRT were repeatedly randomised to notifications over time with outcomes measured after each randomisation [180, 222, 163, 42, 162]. Our aim is to provide evidence of how notifications affect near-term engagement as well as consideration towards further improvement of the push notification policy. The primary objective was to assess if sending a notification at 8 PM increases behavioural engagement (opening the app) in the subsequent hour with *Drink Less*. Secondary objectives included the comparison of two different types of notifications, effect moderation by time and the exploration of effect moderation by a user's context (with context being a user's dynamic state of engagement or habituation). We aimed to understand the role of a notification policy more generally for time-to-disengagement. Additionally, we also aimed to compare three policies on time to disengagement (each policy being the decision rule of delivering notifications employed in one of the three arms). This is a first step in our wider aspiration to optimise the notification policy of *Drink Less*, an aspiration we return to in the discussion section.

10.1 Methods

10.1.1 Trial Design

Our study is a 30-day MRT with two additional parallel arms. Three different notification policies are implemented in the two arms and the MRT, to address secondary objectives. The different policies are (i) a standard policy of sending a daily message of *"Please complete your mood and drinks diary"* sent at 11 AM (ii) the MRT, a random policy which varies the content and sequence of the notifications, and (iii) a no-notification policy, a policy which no notifications are sent. For the secondary objectives, the three policies are referred to as (i) the standard notification policy, (ii) the random notification policy, and (iii) the no-notification policy. Sixty percent of eligible users were randomised to the MRT, and forty percent of eligible users were randomised in equal number to the two parallel arms, either receiving the no notification policy or the standard notification policy, of "Please complete your mood and drinking diary" at 11 AM. For users randomised to the MRT, each user was randomised daily at 8 PM, to receive one of the three options: no notification (with 40% probability), the standard message (with 30% probability), or a notification randomly selected with replacement from a bank of new messages (with 30% probability). Following our MRT protocol [19] and the Consolidated Standards of Reporting Trials (CONSORT) 2010 guidelines [38] we report the primary and some secondary results here.

10.1.2 Participants

The recruitment period ran from 2nd January 2020 to 1st April 2020. *Drink Less* is freely available on the Apple App store, and individuals who downloaded the app during the recruitment period were eligible to participate in the trial if they self-reported a baseline Alcohol Use Disorders Identification Test (AUDIT) score of 8 or above which is indicative of excessive alcohol consumption [5]; resided in the UK; were aged 18 years or over; and reported being interested in drinking less alcohol.

The app prompted eligible users to read the privacy notice (Appendix 2) and participant information sheet (Appendix 3) before proceeding to enrol in the trial. During the informed consent process, users were informed that they could opt out of the trial at any time and that they would receive the standard version of the app if at any time they withdrew their consent. Date of download is defined as the date which the onboarding process is completed by each user. The onboarding process involved users completing a registration section where they completed the AUDIT and sociodemographic assessment, and then received normative feedback (personalised feedback on how their drinking compares with the behaviours of others). It is only after completion of the onboarding process that users were then assessed for eligibility and consequently randomised to one of the three arms. Upon enrolment to the study, we turned the permission function off within the app. This was with the intention to ensure that the participants received the notification policy

they were randomised to. Participants could, however, go into the settings and turn the notification policy off, which is applicable for all apps on the Apple App Store and is beyond the control of any app developers.

10.1.3 Data

Pre-processing of the original usage data was required. The raw engagement data is captured by a series of screen views, comprising of timestamps of when a new screen is opened in the app. Clearing or swiping away the notification does not register as any use [40]. The length of a session is calculated as the difference (in micro-seconds) between the first screen view and the last screen view, with a new session defined after 30 minutes of inactivity between screen views [41]. This method of calculating length of sessions means our measures of the length of time spent on the app are always underestimated, as we do not know how long the user observed the last screen view [86]. We did not impose a threshold on our outcome (in terms of amount or depth of app use), so simply opening the app is measured as engagement. When a user opens *Drink Less*, they are presented with a dashboard with various information about their drinking habits as well as a 'toolbox' of features to access if they want. As such, simply opening the app and viewing the dashboard and toolbox presents an opportunity for users to benefit from engaging with *Drink Less*. All time stamps were appropriately adjusted from Coordinated Universal Time to British Summer Time.

10.1.4 Time-fixed Measures (Baseline)

Time-fixed covariates, measured at baseline were age, sex, type of employment (manual, non-manual, or other) and baseline AUDIT score (0-40) [5, 195]. The AUDIT risk zones were: hazardous (8-15), harmful (16-19) and at risk of alcohol dependence (20-40). The participants self-selected the employment status they identify with for the options they were provided. They were not provided with a definition of the employment type.

Time-Varying Measures

Time-varying engagement measures within the MRT are timestamps of when the user opens the app and the length of time (in seconds) spent on the app. This includes the time-varying variables (i) "did the user open the app before 8PM on day of randomisation? (yes/no)" (ii) "did the user open the app any time after 9PM the day before? (yes/no)". Time-varying covariates, used as part of post-hoc analyses to explore effect moderation, were 'habituation' and 'already engaged'. 'Habituation' was captured using the binary measure "did the user receive a notification the day before? (yes/no)". 'Already engaged' was captured using the binary measure "did the user open the app between 8 PM-9 PM the day before? (yes/no)".

10.1.5 Interventions

The MRT tested two different notification types. This trial tests the existing notification with the message of "Please complete your mood and drinking diary" and a new notification bank of 30 novel messages (Appendix One). The development of the new notification bank was informed by research with *Drink Less* which found that the perceived usefulness of the app (the belief that using the app will help the user achieve their goal(s) and an indicator of users' reflective motivation to engage) was associated with increased engagement for some users. The new bank of notifications was therefore designed (with feedback on the content sought from a group of behavioural scientists) to increase users' reflective motivation to engage with a particular intervention module [172]. All messages contained the phase "(using a particular module in the app) can help you drink less". Examples include "Recording if-then plans can help you drink less" and "Setting a doable goal can help you drink less. Take a moment to set a doable goal". The notification does not lock the users' screen and there is no expiry time to the notification.

10.1.6 Outcomes

The primary outcome is whether the user opened the app (yes/no) in the hour between 8 PM to 9 PM, following the randomisation of receiving a notification at 8 PM. This is a time-varying, binary, near-term measure of engagement. We also defined a post-hoc outcome of whether the user opened

the app between 8 PM on the day of randomisation to 8 PM the following day, to explore the effect over a 24-hour period. Secondary outcomes captured across the three different polices include number of days to disengagement, with disengagement defined as the first day in a period of seven or more consecutive days of no use.

10.1.7 Sample Size

We powered the MRT for the important secondary objective of effect moderation over time, which guarantees at least as much power for the primary objective to detect a marginal effect. Using simulation informed by observational *Drink Less* data [19], we determined that a sample of 1,200 users was sufficient to provide 80% power, with type I error of 5%, to detect effect moderation over time, assuming a marginal notification effect of 2.16 decaying by 0.911 per day since download. We powered the secondary arms, implementing different notification policies, to detect a minimum absolute difference in time-to-disengagement of 10%, assuming 55% disengagement by day 22 under the standard policy compared with 65% under the no-notification policy. To achieve 80% power, with type I error of 5%, we required 372 users to receive each notification policy. This was rounded to 400 to simplify the randomisation process. Overall, we aimed to recruit 1,200 users to the MRT, 400 users to the standard notification policy and 400 users to the no notification policy. Previous download trends revealed, on average, an estimate of at least 33 eligible users per day who downloaded Drink Less and consented to the privacy

notice, and we expected the available recruitment window (2nd January to 1st April 2020) to be sufficient to reach our recruitment target of 2,000 users. However, we fell short of this target of 2,000 users and randomised 598 users in total for three reasons: (i) a large proportion of users (40%) did not give their informed consent to be part of the study, (ii) the number of downloads, particularly for March 2020, was less than predicted, based on 2019 trends, and (iii) extending the recruitment period to achieve the desired sample size was not possible due to the commencement of a prescheduled NIHR-funded RCT [84]. Consequently, the primary objective was sufficiently powered, yet we did not achieve the pre-specified sample size for the secondary objectives of effect moderation over time and time-to-disengagement.

10.1.8 Randomisation

Simple randomisation (unstratified and no blocking) was used. An external engineer generated the randomisation sequence and coded this into the app. Two co-authors (LB and CG) pilot tested the randomisation schedule. To further verify the randomisation process, ten volunteers also participated in a pilot test. The ten volunteers were randomised to the three different arms and asked to record notifications received and use of the app. We confirmed the randomisation process functioned as planned, and all use was correctly captured.

10.1.9 Statistical Methods within the MRT

Descriptive statistics (frequency distributions and measures of central tendency) were used to describe the baseline variables of participants. The primary outcome, within the MRT, was summarised separately for the standard notification, the new notification, and no notification, by the number of person-days where the app was opened between 8 PM to 9 PM then divided by the number of person-days in the MRT, and expressed as a proportion. The near-term effect of the notification on the primary outcome was expressed on the relative risk scale, and pooled over the longitudinal data across all participants, using the Estimator for the Marginal Excursion Effect (EMEE) [44]. The EMEE was developed for estimating causal effects of time-varying treatments with binary outcomes. The EMEE does not require the correct specification of the marginal mean model (i.e., how the time-varying engagement depends on a user's time-varying contexts), providing robustness to highly complex and stochastic engagement patterns. The effect of receiving a push notification, versus not receiving a notification, was estimated overall and then separately for the new message bank and the standard notification. All models from the MRT were adjusted for the continuous variables of age, AUDIT score, days since download, the categorical variables of sex and employment type and the time-varying variables "did the use user the app before 8 PM that day?" and "did the users use the app after 9PM the day before?". The time-varying measures were included to increase the precision of our near-term notification effect, as they are likely highly

correlated with the outcome. The covariates of 'habituation' and 'already engaged' are for the purpose of exploring how these recent states modify the near-term effect of the notification.

10.1.10 Statistical Methods across Arms

Baseline descriptive statistics and measures of use – the median number of sessions per user and the median length of sessions (seconds) – were reported across the three policies. A user was classified as having disengaged at the first day of a period of 7 consecutive days of no use. This outcome was only defined for the first 23 days since follow-up lasted 30 days in total. Survival curves were plotted using the Kaplan Meier estimator [45] and compared using a log-rank test. Due to technical glitches, there was some unanticipated missing categorical baseline data. We report the number of missing values per arm. We used modal imputation for baseline variables. To assess sensitivity of our conclusions to our missing data approach, we imputed data with the second most common value. All analyses were conducted using R v.4.0.5 [212] with the dplyr [226], lubridate [92], gtsummary [203], zoo [236], ForImp [205], and survminer [112] packages.

Ethics

Ethical approval for this study was granted by the University College of London's Departmental Research Ethics Committee (CEHP/2016/556) on October 11, 2019, and The London School of Hygiene and Tropical Medicine

Interventions Research Ethics Committee (17929) on November 27, 2019.

10.2 Results

The anonymised datasets, including data dictionaries, are publicly available here: `https://osf.io/w3szp/`. The code for the EMEE is openly available on GitHub account: `https://github.com/lauren-bell/DrinkLessMRT`.

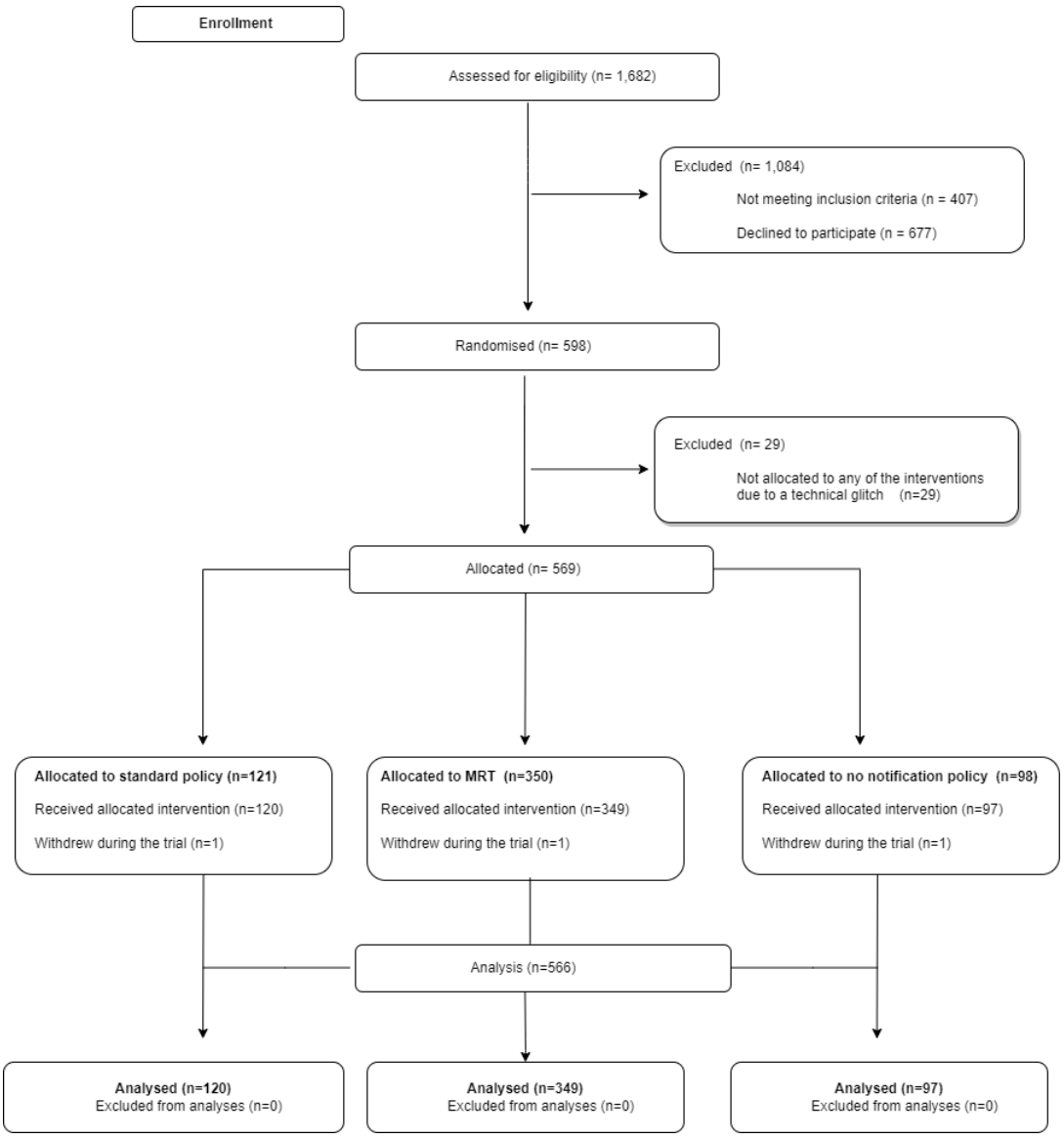


Figure 10.1: Flowchart of Trial

Recruitment

The trial recruitment period ran from 2nd January 2020 to 1st April 2020 (app version 2.0.1). We analysed 566 users in total. Mean age was 44 years

old (SD = 12), with 247 (43.6%) males and 259 (45.8%) females. 353 (62.3%) of users reported being in non-manual employment; 71 (12.5%) in Manual employment; and 82 (14.4%) in Other. 276 (48.8%) reported Hazardous alcohol consumption, 114 (20.1%) reported Harmful and 176 (31.1%) reported to be at risk of alcohol dependence. 386 (68.2%) users disengaged by day 23 or before.

Sex and employment type was not recorded for 60 users (Standard Arm n=5; MRT n=40; No Notification Arm n=5). We used mean modal imputation for these 60 observations of baseline outcomes of sex (to female) and employment type (to non-manual).

		Baseline Summary		
User Characteristics	N	**Standard Arm**, N=120	**MRT**, N=349	**No Notification Arm**, N=97
Age (years), median (IQR)	566	45 (35,55)	43 (31,51)	43 (34, 52)
Sex, n (%)	506			
Male		43 (41%)	155 (50%)	49 (53%)
Female		62 (59%)	154 (50%)	43 (47%)
-Missing-		15	40	5
Employment Type, n(%)	506			
Non-manual		66 (63%)	224 (72%)	63 (68%)
Manual		19 (18%)	37 (12%)	15 (16%)
Other		20 (19%)	48 (16%)	14 (15%)
-Missing-		15	40	5
AUDIT score, n(%)	566			
Hasardous (8-15)		48 (40%)	142 (41%)	49 (51%)
Harmful (16-19)		29 (24%)	84 (24%)	18 (19%)
At risk of alcohol dependence (20-40)		43 (36%)	123 (35%)	30 (31%)

10.2.1 Outcomes and estimation

In the MRT, 349 users were randomised each day for 30 days, resulting in 10,470 measurements for the primary outcome. There were 3,146 (30.0%)

measurements for the new message, 3,112 (29.7%) measurements for the standard notification and 4,212 (40.2%) measurements for no notification. The proportion of the primary outcome occurring (opening the app between 8 PM-9 PM) was: 0.122 for the new message; 0.131 for the standard message; and 0.036 for no message. For the post-hoc 24-hour outcome (from 8 PM to 8 PM the next day), the proportion of opening the app was 0.351 for the new message; 0.342 for the standard message; and 0.280 for no message.

Main Results

Table 10.1: Primary objective – adjusted marginal effect of receiving a notification compared to not receiving a notification.

Notification Type	**Relative Risk (95% confidence interval)**
Pooled notifications (both standard and new)	3.523 (2.918 - 4.255)
Standard notification	3.664 (2.993 - 4.485)
New notification	3.385 (2.774 - 4.131)

Table 10.2 provides the results for the estimate of the near-term effect of the notification on engagement. This demonstrates that on average, the probability of opening Drink Less within the hour of receiving a notification increased 3.52-fold; 95% CI (2.91 to 4.25). The two different notification types have similar effects, with the probability of opening Drink Less within the hour of receiving a standard notification increasing 3.66-fold; 95% CI (2.99 to 4.48) and the probability of opening the Drink Less app within the hour of receiving a new notification increasing by 3.39-fold; 95% CI (2.77 to 4.13).

Table 10.2: Change of marginal treatment effect (Relative Risk) over time (rate of habituation per day)

Notification type	**Relative Risk (95% confidence interval) on day after download**	**Multiplicative change in effect**
Pooled notifications (both standard and new)	3.849 (2.811 - 5.270)	0.993 (0.975 - 1.012)
Standard notification	4.193 (3.004 - 5.854)	0.989 (0.970 - 1.001)
New notification	3.534 (2.536 - 4.924)	0.997 (0.976 - 1.017)

Table 10.3: Summary across three policies

Arm	**Median (IQR) number of sessions**	**Median (IQR) length of sessions (seconds)**
Standard Arm	14 (4, 32)	28 (6, 87)
MRT	10 (3, 42)	28 (7, 86)
No notification Arm	8 (3, 27)	36 (10, 107)

Time to disengagement – Survival Analysis

The median time to disengagement was 11 days for the standard arm, 11 days for the MRT and 7 days for the no notification arm. The number of users who disengaged was 83 in the standard arm, 232 in the MRT and 71 in the no notification arm. The log rank chi-squared test statistic is 1.7 with 2 degrees of freedom and the corresponding p-value is 0.42.

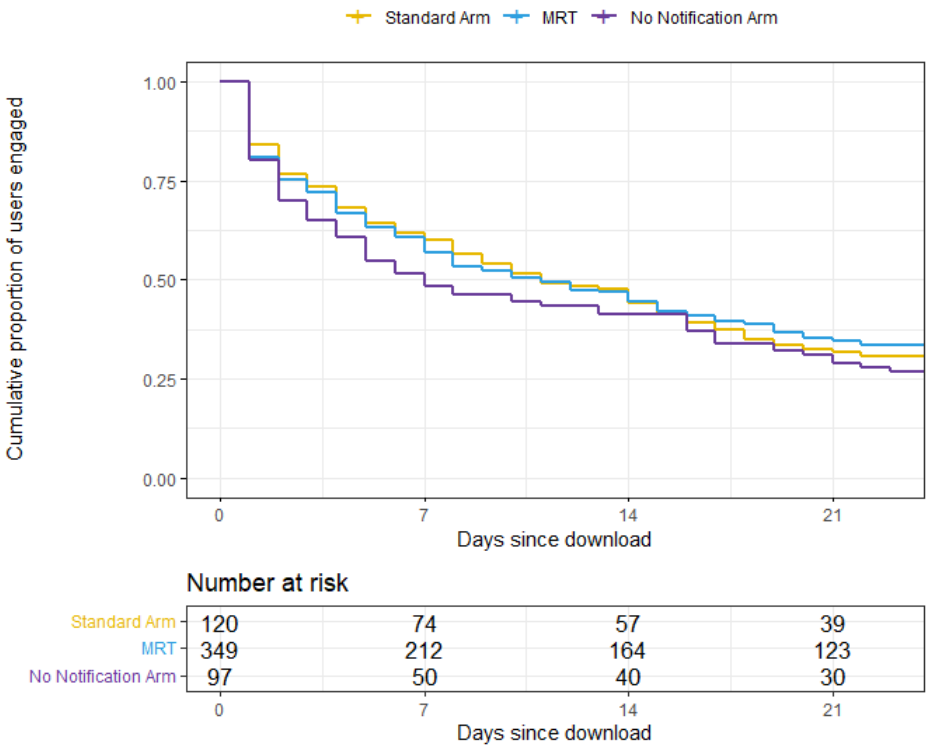


Figure 10.2: Kaplan-Meier plot of Time-to-Disengagement

Ancillary analyses

Table 10.4: Estimated effect and effect moderation (relative risk)

Notification type	**Relative Risk (95% confidence interval)**	**Multiplicative change in effect (95% confidence interval)**
Habituation - "did the user receive a notification the day before?"		
Pooled	3.645 (2.665 - 4.987)	0.946 (0.655 - 1.368)
Standard	3.935 (2.843 - 5.447)	0.889 (0.603 - 1.311)
New Message	3.357 (2.387 - 4.721)	1.013 (0.678 - 1.513)
Already engaged – "did the user open the app between 8PM-9PM the day before?"		
Pooled	3.621 (2.908 - 4.508)	0.875 (0.616 - 1.242)
Standard	3.720 (2.941 - 4.704)	0.962 (0.656 - 1.412)
New Message	3.522 (2.797 - 4.434)	0.800 (0.551 - 1.163)

- Adjusted for the continuous variables of age, AUDIT score, days since download, the categorical variables of sex and employment type and the time-varying variables "did the use user the app before 8PM that day?" and "did the users use the app after 9PM the day before?" and the effect moderation variable of Habituation.
- Multiplicative change in relative risk when effect modifier is positive

Table 10.5: Post-hoc estimated adjusted marginal effect defined over 24 hours (from 8PM to 8PM the next day).

Notification type	**Relative Risk (95% confidence interval)**
Pooled	1.260 (1.187 - 1.337)
Standard	1.245 (1.161 - 1.336)
New Message	1.274 (1.193 - 1.360)

- Adjusted for the continuous variables of age, AUDIT score, days since download, the categorical variables of gender and employment type and the time-varying variables "did the use user the app before 8PM that day?" and "did the users use the app after 9PM the day before?".

10.3 Discussion

10.3.1 Principal Findings

We have shown that, for *Drink Less*, there is a large near-term (3.5-fold) positive effect on engagement. The near-term notification effect for either the standard message type or a message from the new bank have similar effects in increasing engagement in the subsequent hour. Over a 24-hour period, a smaller, significant effect (1.3-fold) remains. We did not detect a significant change in the effect of the notification over time. The effect of receiving a new message, which aims to re-engage users, was non-significantly reduced by 20% if the user was already engaged. Furthermore, the effect of receiving a standard message was non-significantly reduced by 12% if the user received a notification the day before. There was no significant difference in (i) the mean number of days to disengagement, (ii) number of sessions and (iii) length of sessions across the three different notification policies. However, a slightly higher median length of time for a session under the no notification policy was observed. One might hypothesise that behavioural engagement that is

unprompted may include more attentive interest and cognitive investment.

In our study, despite evidence of a large positive notification effect on near-term engagement, an overall policy of sending a fixed daily notification or a random mix of notifications did not lengthen the time to disengagement or increase the amount of engagement during the first 30-days of download. The results of the effect moderation analyses, although requiring confirmation in larger studies, suggest that notifications may be better served as dynamic interventions which adapt to a user's fluctuating patterns of engagement, for example via a policy of sending a notification to users when they are at an increased risk of disengagement, targeting them at that point with a notification intended to increase their perception of the usefulness of the app.

10.3.2 Future research to optimise the notification policy

Our study has demonstrated that, for *Drink Less*, the notification increases near-term engagement. This finding offers the opportunity for behaviour change scientists to directly target the precise momentary states of an individual, to develop and implement dynamic theories for behaviour change with *Drink Less*.

Efforts to maintain or increase engagement through consistent notifications could overburden or annoy a user, resulting in a state of disengagement with the interventions from a previously motivated user[19]. Our findings suggest

that the optimal role of notifications to improve long-term engagement is unlikely to be fixed or random components, but better placed as dynamic components (i.e. varying not randomly but in response to the user's changing state of engagement and habituation). The open question now is when do we programme notifications to be sent, to balance goals of (i) intervening for maximum therapeutic effect, based on a users' internal history with Drink Less and external, environmental factors; and (ii) avoiding states of disengagement due to the burden of unhelpful notifications. To begin to answer this question, we will undertake further modelling of this MRT data, to explore the within- and between- user effect of the notification over time, and the balance of near-term and long-term effects. We will further analyse the data to understand if cue-to-action messages resulted in the task, to determine if the suggested module was engaged with. We imagine a further optimised policy would (i) keep more users in a state of engagement for longer by sending fewer notifications than the policies tested here, (ii) have a higher near-term notification effect, and (iii) ultimately improve the effectiveness of *Drink Less*. A type of machine learning, called reinforcement learning, may be helpful to personalise and optimise the sequence of notifications over time [238]. The available data from our trial can provide a rich source of information to help guide the initial steps (i.e., provide a "warm-start") of the learning process of a reinforcement learning algorithm, to improve engagement, for *Drink Less* or other similar behaviour change apps [238, 233, 132, 133].

10.3.3 Limitations

Our study was sufficiently powered for the primary objective, to detect a near-term notification effect. However, due to not achieving our planned sample size, important secondary objectives of effect moderation over time and time to disengagement between policies were not adequately powered. This resulted in wide confidence intervals and large p-values for the effect moderation analyses, leaving remaining uncertainties about the existence and magnitude of these effects. Larger studies are required to explore these effects. There was missing data for a minority of the baseline values of sex and employment type, though our sensitivity analyses showed that the result was not sensitive to how the missing values were imputed. The values entered for alcohol units consumed as diary entries were deemed too noisy to represent alcohol consumption over time due to bias, extensive missing data and backfilling (i.e. users bulk reporting their drinking outcomes days later). Due to a priority to not overburden users with too many notifications sent within a day, our research does not provide a comparison of the near-term effect of the notification for different times of the day.

10.3.4 Generalisability

The recruitment period was from 2nd January to 1st April 2020, which began with a typical surge in downloads in the new year and ended during the United Kingdom's first COVID-19 lockdown. Such significant changes

in the overall environment during the trial are likely to influence the underlying thoughts, emotions, and behaviours about reducing drinking levels (i.e., people were mainly housebound) [59] and hence impact the patterns of engagement with *Drink Less*. The interpretation of the result is an average over this period only, with most of the recruitment occurring before the wide-spread outbreak in the UK (Appendix 5). We also note the median time to disengagement in the standard policy arm (11-days) is much sooner than our data visualisation cohort experienced (22 days) [18].

10.3.5 Conclusions

We found a large causal effect of sending a notification on near-term engagement. The probability of opening the app in the immediate hour increased 3.5-fold when receiving a notification, compared to not receiving a notification. Notifications are important and effective components of behaviour change apps; however, a policy of sending a fixed daily notification or a randomly chosen series of notifications did not increase the amount of engagement, or length of time to disengagement for users compared to a policy of no notifications. This suggests notifications may better serve users when they are implemented as dynamic components, such as sending a notification to increase the perceived usefulness of the app only when the users' pattern of engagement shows they are at risk of disengaging. Further optimisation of the notification policy is required to achieve an improvement in long-term engagement. The next stage of research is to explore how our findings would

help develop a policy for *Drink Less* to both intervene when a user is likely to benefit from support and keep more users engaged for the first 30-days since download.

10.3.6 Other Information

Registration Registration number and name of trial registry International Registered Report Identifier (IRRID): DERR1-10.2196/18690

10.4 Other Information

10.4.1 Registration

Registration number and name of trial registry

International Registered Report Identifier (IRRID): DERR1-10.2196/18690

10.4.2 Protocol

Where the full protocol can be accessed

The published protocol can be found here:

Bell L, Garnett C, Qian T, Perski O, Potts HWW, Williamson E Notifications to Improve Engagement With an Alcohol Reduction App: Protocol for a Micro-Randomized Trial JMIR Res Protoc 2020;9(8):e18690 doi: 10.2196/18690PMID: 32763878 PMCID: 7442945

10.4.3 Funding

Source of funding and other support (such as supply drugs), role of funders

The development of the *Drink Less* app was funded by the National Institute for Health Research (NIHR) School for Public Health Research, Society for the Study of Addiction, Cancer Research UK, and the UK Centre for Tobacco and Alcohol Studies (UKCTAS). The NIHR School for Public Health

Research is a partnership between the Universities of Sheffield, Bristol, Cambridge, and Imperial; University College London; The London School for Hygiene and Tropical Medicine; LiLaC: a collabo-ration between the Universities of Liverpool and Lancaster; and Fuse—The Centre for Translational Research in Public Health: a collaboration between Newcastle, Durham, Northumbria, Sunderland, and Teesside Universities. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. UKCTAS is part of UK Clinical Research Collaboration, a Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council, and the NIHR under the auspices of the UK Clinical Research Collaboration is gratefully acknowledged.

The authors would like to thank Dr Dave Crane for his important role in the development and factorial screening trial of the *Drink Less* app and the National University of Singapore's Institute of Mathematical Sciences for funding LB and TQ's visit to the program on Statistical Methods for Developing Personalized Mobile Health Interventions. The authors would also like to thank the UCL Tobacco and Alcohol Research Group for their helpful feedback on the draft of this paper.

LB is supported by a PhD studentship funded by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2-R18).

CG is funded by the NIHR (NIHR127651), the NIHR School for Public

Health Research and Cancer Research UK (CRUK: C1417/A22962).

TQ is supported by the National Institute on Alcohol Abuse and Alcoholism of the National Insti-tutes of Health under award number R01AA23187; National Institute on Drug Abuse of the National Institutes of Health under award number P50DA039838; National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54EB020404; and National Cancer Institute of the National Institutes of Health under award number U01CA229437. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

OP is funded by Cancer Research UK (CRUK: C1417/A22962).

HWWP receives funding from Public Health England, the NIHR and NHS England.

EW is supported by the MRC (MR/S01442X/1) and Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), and British Heart Foundation and Wellcome.

10.4.4 Conflicts of Interest

Only authors CG and HWWP have conflicts of interest to declare:

CG is part of the team that developed and are currently evaluating the *Drink Less* app as part of a randomised controlled trial (ISRCTN64052601) and is a paid scientific consultant for the behaviour change and lifestyle organisation 'One Year No Beer'.

HWWP has received additional salary support from Public Health England and NHS England. He has a PhD student who works at and has fees paid by Astra Zeneca, and another who works at Better Points. He has research collaborations with Thrive Therapeutic Software Ltd. and has had one with Six to Start. He has an engagement project in collaboration with Digital-Health.London, Zinc and BMJ Innovations.

10..5 Bank of 30 newly developed messages and their link to the relevant behaviour change module

Table 10.6: Bank of 30 newly developed messages and their link to the relevant behaviour change module.

Message content	**Behaviour change module**
Tracking your drinks and days you don't drink can help you *Drink Less*.	Goal Setting
Did you know that tracking your drinks and days you don't drink can help you *Drink Less*?	Goal Setting
Tracking your drinks and days you don't drink can help you *Drink Less*. Take a moment to track your drinks or a drink-free day.	Goal Setting
Setting a doable goal can help you *Drink Less*.	Action Planning
Did you know that setting a doable goal can help you *Drink Less*?	Action Planning
Setting a doable goal can help you *Drink Less*. Take a moment to set a doable goal.	Action Planning
Tracking your mood after drinking and drink-free days can help you *Drink Less*.	Self-Monitoring and Feedback
Did you know that tracking your mood after drinking and drink-free days can help you *Drink Less*?	Self-Monitoring and Feedback
Tracking your mood after drinking and drink-free days can help you *Drink Less*. Take a moment to track your mood.	Self-Monitoring and Feedback
Tracking your productivity levels after drinking and drink-free days can help you *Drink Less*.	Self-Monitoring and Feedback

Did you know that tracking your productivity levels after drinking and drink-free days can help you *Drink Less*?	Self-Monitoring and Feedback
Tracking your productivity levels after drinking and drink-free days can help you *Drink Less*. Take a moment to track your productivity levels.	Self-Monitoring and Feedback
Tracking your sleep quality after drinking and drink-free days can help you *Drink Less*.	Self-Monitoring and Feedback
Did you know that tracking your sleep quality after drinking and drink-free days can help you *Drink Less*?	Self-Monitoring and Feedback
Tracking your sleep quality after drinking and drink-free days can help you *Drink Less*. Take a moment to track your sleep quality.	Self-Monitoring and Feedback
Tracking how clear-headed you feel after drinking and drink-free days can help you *Drink Less*.	Self-Monitoring and Feedback
Did you know that tracking how clear-headed you feel after drinking and drink-free days can help you *Drink Less*?	Self-Monitoring and Feedback
Tracking how clear-headed you feel after drinking and drink-free days can help you *Drink Less*. Take a moment to track your clear-headedness.	Self-Monitoring and Feedback
Keeping an eye on how your drinking compares with others can help you *Drink Less*.	Normative Feedback
Did you know that keeping an eye on how your drinking compares with others can help you *Drink Less*?	Normative Feedback
Keeping an eye on how your drinking compares with others can help you *Drink Less*. Take a moment to check how your drinking compares with others.	Normative Feedback

Recording if-then plans can help you *Drink Less*.	Action Planning
Did you know that recording if-then plans can help you *Drink Less*?	Action Planning
Recording if-then plans can help you *Drink Less*. Take a moment to record an if-then plan.	Action Planning
Keeping an eye on which if-then plan has and hasn't worked can help you *Drink Less*.	Action Planning
Did you know that keeping an eye on which if-then plan has and hasn't worked can help you *Drink Less*?	Action Planning
Keeping an eye on which if-then plan has and hasn't worked can help you *Drink Less*. Take a moment to check your if-then plans.	Action Planning
Playing the "yes please, no thanks" game can help you drink less.	Cognitive Bias Re-training
Did you know that playing the "yes please, no thanks" game can help you *Drink Less*?	Cognitive Bias Re-training
Playing the "yes please, no thanks" game can help you *Drink Less*. Take a moment to play the game.	Cognitive Bias Re-training

Appendix 10.A Privacy notice

Privacy notice

The data we collect is anonymous and will only be used for academic research conducted by University College London to develop improved ways of helping people to *Drink Less*.

We are currently testing how the push notifications you receive affect how you use the app. In this research, you may be randomized to receive either no messages, the same message, or different messages over time.

The Tobacco and Alcohol Research Group at University College London (UCL) is collecting the data and processes users' data for the following purposes:

- To help users *Drink Less*
- To improve the smartphone application
- To conduct research and write publications that add to the scientific literature
- To communicate with users

The data collected includes that which the user voluntarily enters or provides when using the *Drink Less* app (eg, gender, age, country, job type, alcohol consumption) and app usage. App versions up to and including 1.0.10 col-

lected personal data—email addresses—which was entered voluntarily.

Those personal data will be stored for three years.

Processing of users' data is necessary for the performance of a task carried out in the public interest. Health data on alcohol consumption is necessary for scientific research purposes and is in accordance with safeguards.

Research governance within UCL ensures that data is:

- Necessary to support research
- Only used to support legitimate research activities that are considered to be in the public interest
- Safeguarded/protected

Disclosure:

The Tobacco and Alcohol Research Group at UCL will not share user data with any third parties.

Right of access:

As the data collected is anonymous, it is not possible to link any data we hold with an individual user. For versions up to and including 1.0.10 of the app, email addresses were collected from some users. If a user downloaded one of these app versions and entered their email address, then the Tobacco and Alcohol Research Group at UCL can share the data held on the user. Up

to three years after downloading the app, users can request this by emailing support@drinklessalcohol.com, along with the email address entered during the registration process. The Tobacco and Alcohol Research Group at UCL will then send users the data that it holds on them within one calendar month of receipt.

If you have any questions about this privacy notice, please contact us at support@drinklessalcohol.com

Terms and conditions

All data will be stored securely and in line with our privacy notice. You are not obliged to have your data used for academic research, and you should not feel coerced. If you choose to withdraw, you may do so without a disadvantage to yourself and without any obligation to give a reason. To withdraw, please go to the Help tab of the app and choose "Opt out of the study."

Please feel free to ask us any questions on support@drinklessalcohol.com

Consent

By consenting to this Privacy Notice, you are explicitly giving the Tobacco and Alcohol Research Group at UCL permission to process your data for the purposes specified.

You may withdraw consent at any time by going to the Help tab of the app and choosing "Opt out of the study."

I consent to the use of my data, as explained by the Privacy Notice and Terms and Conditions.

[Yes, I agree]

[No, I disagree]

Information Sheet

[Screen2] Information sheet for a research project

You are being invited to take part in a research project. Before you decided to take part, it is important for you to understand why the research is being done and what participation will involve. The information here will try and answer any questions you might have about the study but contact us at support@drinklessalcohol.com if there is anything else you would like to know. Please read the following carefully and discuss it with others if you wish.

Who is conducting the research?

We are a team of researchers at University College London and London School of Hygiene and Tropical Medicine.

Why are we doing this research?

We want *Drink Less* to be an app that you want to use. There are features in the app which may or may not encourage you to carry on using the app. We want to know how to improve these features. We are only interested in how you use the app during the first 30 days.

Why have I been chosen?

You are eligible to take part in this study if you are aged 18 years old or older, have indicated that you are interested in using the app for drinking

less alcohol, and have a score of 8 or more on the test to identify alcohol use disorders, as we are particularly keen to support this group in continuing to use the app. Most importantly, you decide to take part or not.

What will happen if I choose to take part?

If you choose to take part in this study, the reminders that *Drink Less* sends you may have different wording compared with people who choose not to take part. Otherwise, your experience with the app will not be different from other people.

Could there be problems for me if I take part?

We do not anticipate any problems caused by taking part in this study.

What data will you collect about me?

We will not at any time access any personal information about you, such as your name, address, email address. We will collect data about your age, gender, occupation type, and your alcohol consumption when you download the app. For the following 30 days, we will collect data about how you use the app, such as did you use the app or how long you spent on the app. We cannot identify you in this study, and the app will not provide us with any other information available on your phone.

What will happen to my data at the end of the research?

At the end of the study, we will make a dataset available to the public here:

https://osf.io/q8mua. It will not be possible for anyone to identify you in the dataset. The dataset will only contain your general characteristics (age, gender, occupation type, baseline alcohol consumption) and how you used the app for 30 days, such as did you use the app or how long you spent on the app. The content of the diaries will not be made available to us or anyone else. If you have any questions, please feel free to ask Professor Susan Michie on support@drinklessalcohol.com

Consent

You are about to consent to a scientific study that examines how people use this app.

By taking part, you agree that you have read and understood the information about the study.

By consenting to this Privacy Notice, you are explicitly giving the Tobacco and Alcohol Research Group at UCL permission to process your data for the purposes specified.

I consent to my data being fully anonymized and made available in the public domain via a data repository. **Tick Box** [Yes, I agree] [No, I disagree]

I understand that it is not possible to identify me at any stage in this research, or from any information made public. **Tick Box** [Yes, I agree] [No, I disagree]

I understand I may withdraw consent at any time by going to the Help tab

of the app and choosing "Opt out of the study." **Tick Box** [Yes, I agree] [No, I disagree]

I consent to the use of my data, as explained by the Privacy Notice and Terms and Conditions.

[Yes, I agree]

[No, I disagree]

10.A.1 Sample size calculation

Based on findings from the exploratory analyses of *Drink Less* users (manuscript submitted), we create generative models with a varying magnitude of treatment effect moderation with a few plausible values of marginal treatment effect. In particular, the exploratory analysis suggested that a crude estimate of the marginal treatment effect is 2.16 (in terms of relative risk), and the effect is modified by a factor of 0.911 for each additional day since download. Therefore, in the generative models, we let the marginal treatment effect to take values from 2.16, 1.80, and 1.50 (in terms of relative risk). For each marginal treatment effect value, we consider the magnitude of effect moderation by day-in-study (range –0.03 to –0.012) (in terms of log relative risk, assuming a log-linear model; eg, effect moderation–0.015 means that for each additional day in the study, the treatment effect decreases by a factor of exp(–0.015) = 0.985, and at the end of the study (day 30) the treatment effect will be exp(–0.015 $\times$ 29) = 0.647 of the treatment effect on the first day of the study). Figure 2 shows the result of the simulation-based sample size calculation, where the power to detect moderation for each sample size, marginal treatment effect, and moderation combination is calculated from 1000 simulations, where the critical value of the t-test is chosen to ensure 0.05 Type I error control.

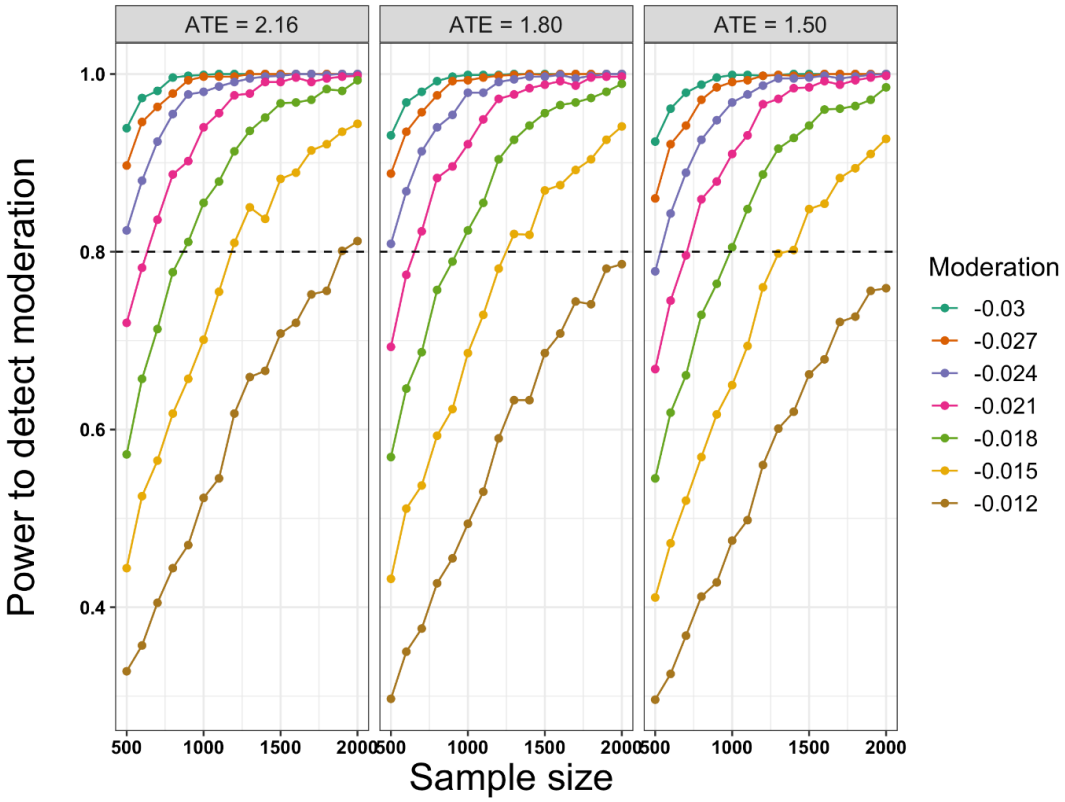


Figure 10.3: Result of the simulation-based sample size calculation.

"ATE" (average treatment effect) denotes the marginal treatment effect averaged over 30 days (relative risk scale). "Moderation" denotes the treatment effect moderation by day-in-study (on log relative risk scale, assuming a log-linear model; eg, effect moderation being –0.03 means that for each additional day in the study, the treatment effect decreases by a factor of exp(–0.03) = 0.97).

Simulations indicate that to ensure roughly 80% power to detect a –0.015 moderation (yellow curve in Figure 2) or larger moderation in magnitude,

a sample size of 1200 is sufficient. Comparison across the three panels in Figure 2 shows that the power to detect moderation is relatively insensitive to different marginal treatment effect values. Because the sample size required to detect a nonzero marginal treatment effect is drastically smaller than the sample size required to detect moderation, we have close to 100% power to detect a nonzero marginal treatment effect under all of the sample sizes, marginal treatment effect, and moderation combinations considered in the simulation (results not shown). Therefore, we set the target sample size of the MRT arm to be 1200. In the simulation-based sample size calculation, we consider the following generative model. The binary outcome (whether or not the user logged into the app or not between the hour of 8PM to 9PM) on day t is generated from a Bernoulli distribution with success probability

$$\boldsymbol{\exp\left\{\alpha_0 + \alpha_1 t + A_t\left(\beta_0 + \beta_1 t\right)\right\}}$$

where $\boldsymbol{t}$ denotes day-in-study $\boldsymbol{(1, 2, \ldots, 30)}$, and $\boldsymbol{A_t}$ is the binary indicator of whether a notification is sent on day $\boldsymbol{t}$, with $\boldsymbol{0.6}$ probability to be 1 and $\boldsymbol{0.4}$ probability to be 0 . Following (Qian—et al., 2020) the marginal excursion effect averaged over time on relative risk scale (ie, "average treatment effect," or ATE) is defined as

$$\text{ATE } = \frac{\sum_{t=1}^{30} \exp\left(\alpha_0 + \alpha_1 t + \beta_0 + \beta_1 t\right)}{\sum_{t=1}^{30} \exp\left(\alpha_0 + \alpha_1 t\right)}.$$

Using the result of the prior cohort data, in the generative model, we fix the

success probability of the binary outcome under no treatment on day 1 to be $\mathbf{7.6\%}$, and we fix the success probability of the binary outcome under no treatment on day 30 to be $\mathbf{2.2\%}$. The effect moderation by day-in-study on log relative risk scale is $\boldsymbol{\beta_1}$, which we vary between $\mathbf{-0.030}$ and $\mathbf{-0.012}$. We also vary ATE among three values: $\mathbf{2.16}$ (the crude estimate from the cohort data analysis), and two conservative values $\mathbf{1.80}$ and 1.50. Hence, for a given (ATE, $\boldsymbol{\beta_1}$) pair and the fixed success probability under no treatment on day 1 and day 30 , we can compute the values of $\boldsymbol{\alpha_0, \alpha_1, \beta_0}$. The combination $\boldsymbol{(\alpha_0, \alpha_1, \beta_0, \beta_1)}$ determines the generative model.

Under each generative model, for a fixed sample size $\mathbf{(500, 600, \ldots, 2000)}$, we simulate 1,000 data sets and calculate the $\boldsymbol{t}$-statistic for testing $\boldsymbol{H_0 : \beta_1 = 0}$ using the EMEE estimator in Equation 10 (Qian et al., 2020)of by setting $\boldsymbol{S_t = (1, t), g\left(H_t\right) = (1, t), \tilde{p}_t\left(S_t\right) = 0.6}$, and with a small modification in that the factor $\mathbf{exp}\left(-\boldsymbol{A_t S_t^T \beta}\right)$ is replaced by $\mathbf{exp}\left\{-\left(\boldsymbol{A_t - 0.6}\right)\boldsymbol{S_t^T \beta}\right\}$. The $\boldsymbol{t}$-statistic is formed by dividing the estimator of $\boldsymbol{\beta_1}$ by its estimated standard error, and $\boldsymbol{H_0 : \beta_1 = 0}$ is rejected if the absolute value of the $\boldsymbol{t}$-statistic is larger than 1.96. This ensures $\mathbf{0.05}$ Type I error control. Power under each setting is calculated as the proportion of simulations where $\boldsymbol{H_0}$ is rejected.

10.A.2 Recruitment plots

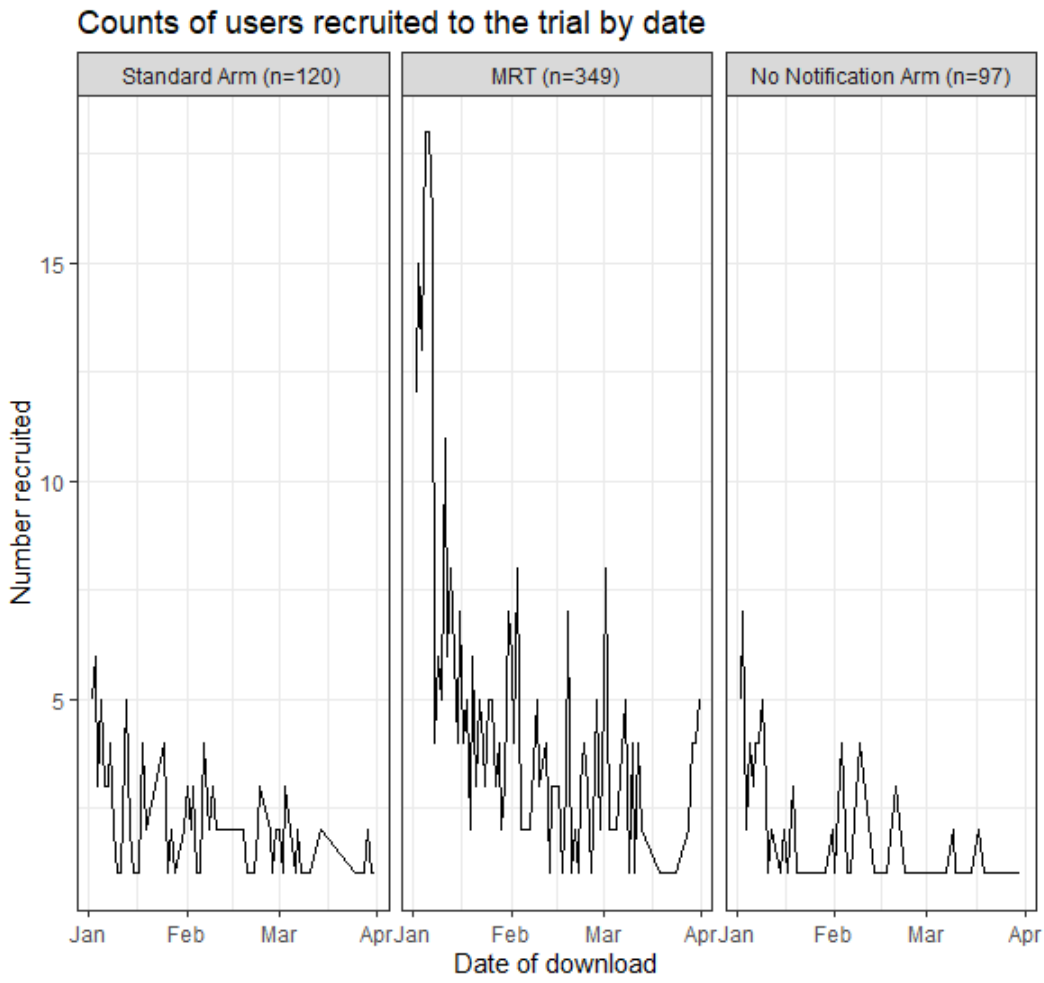


Figure 10.4: Recruitment plots over time

Chapter 11

Discussion

11.1 Summary of thesis

To conclude this thesis, I will summarise my work, review the key discoveries and consider the new research questions and future directions.

I first began by describing two digital therapeutic apps which sought regulatory approval as digital therapies, and considered the differences in the types of evidence evaluated for regulatory approval. This demonstrated the challenges regulatory bodies and behaviour change app developers face with gaps in knowledge of evidence, when evaluation methods do not fit the agile and dynamic nature of behaviour change apps. I then explored the different types of evaluation methods, particularly different types of randomised trial

designs in the literature, through a systematic review.

I began my work with *Drink Less* by exploring patterns of use through simple visualisations. This work suggested that the fixed daily notification policy worked well for some users for maintaining engagement, but not for other users. I concluded that optimising the notification policy is a worthwhile investment, and the future goals were established to (i) determine the causal effect of the notification and (ii) consider tailoring the notification policy to the context of individual users over time. A Micro-Randomised Trial (MRT) would allow us to achieve both goals.

I then designed an MRT, which was complemented with two parallel arms (i.e. a no-notification policy, a random notification policy at 8pm and a standard, fixed notification policy for 11am). I wanted to build a more effective and less burdensome notification policy, with the results of the MRT to inform the optimisation of the content and sequence of the notification policy. The objectives were to understand (i) if the notification had a marginal, near-term effect, (ii) if this effect changes over time or depends on other time-varying states, and (iii) if there is a long-term effect, along with a near-term effect. I also established disengagement as a state, and set the trial to explore trajectories of reaching this state under three different policies.

I then expanded on the work of Qian et. al. [180], by exploring the properties of the EMEE as an estimator for estimating the CEE from MRTs. This work extends the findings of Qian [180], which shows that the EMEE provides con-

sistent estimates for the marginal near-term effect when the adjusted working model omits or does not correctly specify the true relationship between an effect moderator and the outcome under no treatment. The results confirm that GEE with an independence correlation matrix is not guaranteed to be a consistent estimator. The benefits of the EMEE is that it provides a coherent approach to estimate both the primary and secondary objectives for the *Drink Less* MRT.

The results of the MRT showed a large, near-term effect of the notification, but highlighted that a notification policy which does not adapt to the time-varying recent state of engagement does not perform better than no-notification policy. I now hypothesise that a notification policy which only sends notifications to encourage engagement, when users are at a higher risk of disengagement, would increase the effectiveness of the near-term notification while reducing burden, and this may ultimately keep users engaged for longer.

The MRT was analysed using a novel estimator, the Estimator for the Marginal Excursion Effect (EMEE), developed by Qian [180] and discussed in *Chapter Eight*. As discussed, when there is a highly stochastic, complex outcome under the null, the conditional effects are biased for a Sequential Conditional Mean Model (SCMM), when modelled by a Generalised Estimating Equation (GEE) with an independence working correlation matrix, but the EMEE provided consistent results for the Causal Excursion Effect (CEE) for both

the primary and secondary objectives.

Finally, I now conclude with some thoughts towards future research, and future further analysis of the MRT data, to potentially consider and build a warm-start policy for reinforcement learning, developed from the MRT for *Drink Less*, with a constrained contextual bandit in mind. This would be with the goal to both allow (i) personalisation to occur within *Drink Less* and (ii) statements of causal inferences to be made.

The key discovery of my work to the field of engagement in digital health, is how the notification policy impacts engagement with *Drink Less*. Notifications can be useful tools to encourage and direct in-the-moment engagement. However, repeatedly delivering notifications at times which are unhelpful is likely to induce habituation and result in disengagement. Strategies for the notification policy which are myopic (i.e. a policy continues to send notifications to individuals who were already engaged and may have used the app anyway) may result in habituation and result in users to disengage sooner.

Overall, this work contributes to the growing scientific literature that places optimising behaviour change apps to be dynamic health interventions which exist within dynamic environments and individual states. Behaviour change apps can evolve into adaptive interventions, to be more effective for different people with different needs. A better appreciation of the importance of this understanding could significantly change the way apps are developed in the future, with fit-of-purpose evaluation methods that match the agile, dynamic

nature of such interventions.

11.2 The big picture

We live in a world where the most lucrative (and arguably divisive) corporations currently create their profit from personalised algorithms, and these algorithms aim to change purchasing behavior through engagement with their service. Learning algorithms, such as multi-armed bandits and reinforcement learning, are the tools which drive such personalisation in search engines and online advertisement [198, 165, 154, 230]. Monitors and wearable sensors for health will increasingly become more common in our homes [169, 68], and continuous monitoring of health measures could give rise to digital dynamic treatment regimes, with the potential to tailor, in real time, the delivery of treatments, to the context of personal health information gathered over time [176].

In order for personalised behaviour change apps be effective in our health system, there are many important urgent challenges to address [140]. One important challenge is for the intervention to continuously improve through personalisation, while allowing for statements of efficacy to be made.

11.3 Summary of the key conclusions

A key conclusion from this work is that a notification policy within a digital health app is a powerful tool to create in the moment engagement, yet the tailoring to an individual's recent state of engagement and behaviour, environment and past treatments is likely required, to optimise longer-term engagement.

What is a helpful notification for some users may not be helpful for others, or a user may find the same notification helpful in a particular environment or contextual state, but unhelpful in a different contextual state. In this thesis, I argue that sending a push notification, to encourage engagement, when the user is already engaged, is likely to result in delayed negative effects. The development of a notification policy should avoid being myopic, that is only consider the immediate effect when determining if to send more notifications over time.

The cumulative sequence of unnecessary, unhelpful notifications, sent to an individual who is already motivated and engaged (and may respond to the notifications initially) will eventually result in harm in the long term, where harm in this context is defined as burden that induces poor perceptions of the usefulness of the app. This burden will ultimately lead to states of disengagement and render the intervention ineffective for a person who was motivated and engaged with the app at the beginning.

How the effect of the notification changes due to recent contextual states gives a direction and signal for how to better optimise the content and sequence of the notification policy. Optimal development may be better served by considering notifications as strategies to reignite engagement, and sent when the risk of disengagement is increasing for an individual, rather than fixed components to maintain engagement. I now address the open questions which remain.

11.4 Potential future work with *Drink Less*

In this section, I now consider how we could proceed to further develop *Drink Less* into a JITAI or adaptive behaviour change app. I aim to create a notification policy which sends the right notification to the right users at the right moment. To do this we need to better understand how the notification effect is likely to change within and between people over time. This would extend beyond the results reported in the main analysis of the MRT, which represent population average marginal effects.

11.4.1 Sequential decision making

I believe the first steps would be to better understand the between and within-person effects within the notification, and any effects from the cumulative sequence of notifications.

As previously highlighted, a key methodological challenge is to allow the continuous, iterative development of the learning algorithm while facilitating statements of causal effects. Powered constrained bandits offer a solution to this challenge, through allowing the notification policy to be tailored to contextual states of an individual, while guaranteeing sufficient power to estimate the marginal effect of the notification.

Another path for future research is the development of a warm-start policy from the *Drink Less* MRT data. This would be important because poor initialisation of a reinforcement learning algorithm could be an important cause of failure to engage users. To begin, we could consider what a good reinforcement learning algorithm would look like, and consider what information is required to help the reinforcement learning algorithm achieve a good performing initialisation.

Below I briefly describe reinforcement learning, powered constrained bandits and warm-start policy.

11.4.2 Warm-Start policies for reinforcement learning

Reinforcement Learning is a type of machine learning which aims to optimise sequential decision making by a agent (i.e. a JITAI) over time. History is now conceptualised as *states* (i.e. time-varying covariates), *actions* (i.e. time-varying treatment) and *reward* (i.e. time-varying outcomes). The reinforcement learning goal is to learn how to transition between *states* and

actions to maximise the total expected reward [45, 63].

Reinforcement Learning agents can include one or more of the following components.

Policy *How the agent behaves*

Value *How good it is to be in each state*

Model *How the agent represents the environment*

Due to the need for behaviour change apps to keep users engaged from the beginning, the poor initialisation of reinforcement learning should be avoided. This could be achieved through a warm-start policy. Liao [133] listed four characteristics a good reinforcement learning algorithm should consider, which are:

- Adjust for the longer term effects of the current action
- Learn quickly and accommodate noisy data
- Accommodate some model mis-specification and non-stationarity
- Select actions in a way that after the study is over, secondary data analyses are feasible (i.e. causal statements can be made).

11.4.3 Powered constrained bandits

The inspiration for considering Powered Constrained Bandits came from the paper: *Yao, J., Brunskill, E., Pan, W., Murphy, S., & Doshi-Velez, F. (2021, October). Power Constrained Bandits. In Machine Learning for Healthcare Conference (pp. 209-259).*

In this research, Yao et al [233] develop a meta-algorithm which both guarantees sufficient power to detect a marginal effect, while allowing personalisation to occur through contextual bandits. The issues addressed includes minimising regret, which is the difference between the optimal cumulative outcomes for a particular user and the expected outcomes of that policy, while retaining enough power for robust causal effects statements to be made.

In the context of future research for *Drink Less*, a potential next step is to consider creating a **warm-start policy** for a **powered constrained bandit**.

11.5 Conclusion

In summary, this thesis highlights the usefulness of a MRT to better understand the causal effects of a notification on engagement and further pathways of optimising the sequence of the notifications. At a high level, I hope this thesis adds values to the scientific digital health community of how new methods can be implemented and the useful insight they provide. I also hope the

thesis brings awareness of the added benefit behaviour change apps can offer to a user, when their development recognises the way which they can become adaptive and personalised interventions.

Appendix A

Key characteristics of trial designs

A.0.1 Two-arm (simple) RCT

In a two-arm RCT, patients are randomised once to treatments A or B. The treatments are the same for all patients within each treatment group. Randomised patients are followed up over a set time period and patient outcomes collected at the end of this period. The average difference of outcomes between the two treatment groups is estimated as the treatment effect, and determined to be statistically significant or not. Analysis methods are often generalised linear models. The common aim of a simple RCT is to determine if one treatment is more efficacious or effective than another. The control

treatment can be treatment as usual or a placebo. An example of a mobile health app evaluated in a simple RCT is the SMART-PD trial [124]. The main objective of the trial was to determine if the use of a smartphone and web applications to promote self-management would increase treatment adherence. Patients with (i) a diagnosis of Parkinson's Disease, (ii) who had access to a smartphone, tablet or internet connection, and (iii) who required a change to medication regime were recruited to the trial. The control arm was 'treatment as usual' and the treatment period was 16 weeks. The primary outcome was responses to Morisky Medication Adherence Scale (MMAS-8) measured at the end of treatment period of 16 weeks. A generalised linear model compared the average of the outcomes, adjusting for baseline MMAS-18 score and covariates, between treatment arms.

The research questions well suited to RCTs include when interventions are clearly defined and static, and a suitable control arm exists. Unblinding of the outcome data cannot be undertaking during the trial, nor can changes to the interventions. The within and between patient treatment effects cannot be understood.

A.0.2 A/B testing

Within online website development, simple randomised controlled trials are called A/B testing [113, 232]. Participants are randomised to receive two different versions of a website. A common outcome in A/B testing is number

of purchases made or clicks to online advertisements, known as the conversion rate. When more than two websites are compared, the experiment is known as A/B/n tests.

An example of A/B testing is the online double-blind randomised controlled field experiment of an educational website Math Garden [196]. School children visiting the website were randomised to two different versions of an educational programme including mathematical problem solving exercises. Website A was the educational programme used in practise and Website B included a new mechanism which delayed the options to skip solving more difficult mathematical problems. Exerted effort was the outcome. The results showed an increase in the average exerted effort of students who used Website B, compared to the Website A.

A.0.3 Cross-Over Randomised Trials

Patients are randomised to two or more treatments and serve as their own control. For a two-arm cross over randomised trial of treatment A and B, patients are randomised to receive either A then B, or treatment B then A. More than two treatments can be assessed, with k-arm cross over trial designs as k number of treatments.

An example of a cross-over trial includes the teleCRAFT randomised crossover trial, which compared telemonitoring to best practice clinical care for patients with chronic lung disease [48]. The primary outcome was time to first hos-

pital admission for acute exacerbation. The number of patients in the trial was 72 patients, with 61 patients of outcome data on both treatments and 11 patients had some outcomes missing. Paired t-tests were used to compare outcomes between the arms.

This design allows between and within patient treatment effects to be understood. For some treatments, a positive correlation of treatment outcomes within patients is expected, resulting in more statistical power gained compared to a simple RCT with the same sample size. Crossover trials are suited to studying short-term outcomes of chronic diseases or stable processes. Another consideration is 'carry over' effect, in that the sequence of treatments may influence the treatment effect of subsequent treatments [202].A washout period can be implemented between alternating treatments to allow previous treatment effects to subside.

A.0.4 Single Case Experimental Designs

Single-case experimental designs include a family of methods in which participants serve as their own control [62]. This term is commonly found in behavioural research trials, in which the intervention and control alternate over time with outcomes measured for each treatment period. Single case designs have more than one participant, such that the treatment effect is measured within and between participants [61].

It is a misconception that single-case experimental designs are based on

one participant. Types of single-case experimental designs for two treatments A and B include; Reversal (ABA) – when baseline measurements are taken, treatment is implemented and then removed; Multiple-Baseline (AB) – trial commences with baseline, then switching to treatment B is randomly staggered over time across participants; Alternating treatment (ABABAB.....AB) – baseline and multiple treatments are rapidly alternating during the study; Changing Criteria – Treatment adapts to the observed patient behavioural change of the treatment; Combined – Combination of all elements above. An example of single-case experimental design includes the smartphone video-based Cognitive Behavioural Therapy (CBT) for obese patients with an eating disorder [2]. This was a controlled single-case multiple baseline design to understand if a smartphone video-based application increased number of participant's daily meals (as opposed to binge episodes). Five patients were randomised to baseline time lengths between 14 days to 35 days. The total study time length was 55 days for each patient, such that treatment period ranged from 20 to 41 days. In this study, statistical analysis was performed with randomisation tests, non-overlap of pairs calculation.

Single-case experimental designs have the same benefits to randomised cross-over designs, such that within and between patient variability can be understood. The feasibility and acceptance of removing and providing treatment, as well as blinding, needs to be considered for designs.

A.0.5 Series of N-of-1 or N-of-Many

Series of N-of-1 are multiple crossover trials in which both individual and group treatment effects can be summarised [134, 100]. These study designs are ultimately the same as single-case experimental designs, where individuals act as their own control and are randomly assigned sequences of treatments throughout the trial. The estimated population treatment effect is measured as the average difference between outcomes, where individuals can also receive summaries of their own treatments and outcomes.

An example of a series of randomised controlled N-of-1 trials is the StatinWISE trial, which aims to understand if muscle symptoms during statin use are caused by statins in individuals [100]. Patients are randomised to a sequence of six treatment periods, receiving atorvastatin or placebo. The primary analysis will group treatments to test there is an effect on muscle symptoms between treatments, and also provide with patient their own numerical and graphical summaries of treatment effects to individuals.

Series of N-of-1 study design is endorsed when treatment effect is considered heterogeneous across the population, as they enable the between and within treatment effects to be understood. Again they are suited to conditions which are stable and chronic, and there are minimal carry-over effects between treatments.

A.0.6 Sequential Multiple Assignment Randomised Trials

The Sequential Multiple Assignment Randomised Trial (SMART) is a study designs which informs the development of adaptive interventions. Adaptive interventions, also known as dynamic treatment regimes or treatment policies, are a sequence of decision rules where treatment changes over time from the initial responses. The change of treatment is determined by the tailoring variable, which is the patient's measured responses to first line treatments that are part of the adaptive intervention [54].

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. Patients can be re-randomised to one of several intervention options and time points during the trial. The main objective in a SMART is to develop adaptive interventions. Guidance suggests that evaluation of efficacy or effectiveness of the adaptive intervention developed in the SMART is conducted in a subsequent Sequential Multiple Assignment Randomised Trial (SMART).

An example of a SMART includes an adaptive internet-based stress management program called My Health CheckUp [125]. Patients with moderate stress (determined by the Depression, Anxiety and Stress Scale) are recruited and randomised to one of two initial treatments A or B. Treatment A is

self-directed use of My Health CheckUp. Treatment B is minimal guided My Health CheckUp with additional support through weekly telephone lay coaching.

Total treatment duration of the adaptive intervention is 12 weeks, however patient outcomes will be recorded at 6 weeks and defined as non-responders if their stress score did not improve from baseline or remains above the cut-off for mild stress. All non-responders will re-randomised to either (C) continue first stage treatment they were randomised to (that is treatment A or B), or (D) to receive professionally-led motivational interviewing. Responders, that is patients who score below the cut-off for mild stress at 6 weeks, will not be re-randomised and continue first stage treatment, either A or B, for the remaining 6 weeks.

Good adaptive interventions rely on the appropriateness of changing the treatment to a tailoring variable. Dichotomisation of the tailoring variable, such that grouping patients into responders and non-responder to dictate the change of treatment, needs careful consideration.

A.0.7 Micro-Randomised Trials

The Micro-Randomised Trial is a study design which randomly assigns components within an app, such as personalised notifications, repeatedly to participants during a trial. App features tested in a micro-randomised trial are known as 'Just-in-time adaptive interventions' in mobile health research or

push notifications in app development industry. An example of a micro-randomised trial includes the app HeartSteps [117]. HeartSteps is made up of two intervention components (i) daily activity planning and (ii) push notifications of contextually-relevant suggestions of physical activity. There are five decision points during the day when push notifications can be randomised to participants. Push notification were tailored to participant's location, weather, day of the week and time of the day. Randomisation was fixed, such that randomisation was independent of prior randomisations and participant's individual or group responses delivered previously in the study. Analysis was performed by general estimation equation.14, 15 This analysis method has decision points nested within participants and models the outcome within-participant correlation across time. The method incorporates the sequential randomisation to estimate causal treatment effects. Small sample corrections were applied to the critical value of the test statistic to adjust the Type I and Type II errors. Klasnja describes micro-randomisation as a form of a sequential factorial design [117]. MRTs differ from single-case experimental designs by measuring how a component at a set time or context is most efficacious through understanding effect moderation. Simple RCTs do not articulate decision points for components or examine moderators of observed effects.

A.0.8 Multi-Phase Optimisation Strategy Framework

Multi-phase Optimisation Strategy Framework (MOST) is a framework to develop complex interventions made up of multiple components.12 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 screening phase sees the development of a 'first draft' intervention through testing the efficacy of individual intervention components. This is conducted through a between subject randomised factorial (either full or fractional) study, which participants are randomised once to a version of the complex intervention with different module components either present or removed. The refining phase takes the complex intervention, developed from the screening phase, and explores how to optimise the selected individual components, selected from the screening phase, in terms of doses or personalisation. This refining phase is also undertaken with between subject randomised factorial designs. The final confirming phase sees the complex intervention developed from the first two phases evaluated in a RCT.

An example of an app that was development with the MOST framework is *Drink Less* [85, 88, 84]. During the factorial design, five modules were assessed: self-monitoring and feedback; action planning; normative feedback; cognitive bias re-training; and identity change. In the refining phase, each module was developed as a 'high' and 'low' version. For example, the high

version of the self-monitoring and feedback module included info-graphics of calories consumed from alcohol, mood diary and feedback on progress. The low version was only the ability to record number of drinks consumed. The app with all five interventions set to low version was considered the default control. This framework shows how different components within a complex intervention perform, and can be outcomes of acceptability, usability and effectiveness. By randomising patients once to a treatment, the within and between patient treatment effects cannot be understood. With many treatment arms compared to one another, adjustments to the type I error rate may be considered.

A.0.9 Multi-Arm Multi-Stage Trials

Multi-arm Multi-stage (MAMS) trial is a study design in which several treatments, or various combination of treatments, are compared to one control treatment. The trial incorporates interim assessments in which recruitment to a treatment arm is discontinued if deemed to be under-performing. The primary outcome across interim and finals stages can vary. MAMS are adaptive study designs, based on group sequential methods which allow multiple looks at the data during the trial and decisions based on critical values. A trial without the interim analysis is considered a Multi-Arm (MA) trial.

Examples of a MAMS trial the STAMPEDE trial, which assessed multiple therapies for advanced or metastatic prostate cancer [211]. The STAMPEDE

trial began with a control arm and five treatment arms, with a 2:1:1:1:1:1 control to treatment randomisation ratio allocation. The first stage outcomes were based on safety and feasibility objectives, which resulting recruitment stopped to two treatment arms. The second stage was based on failure free survival, and the final analysis is based on overall survival. Because the design tests multiple hypotheses, this multiple testing influences the rate of false positive findings. Guidance for MAMS recommends prioritising the control for the family-wise error rate, that is the probability of a false-positive with multiple testing [224].

MAMS are a more efficient method, in both duration and number of patients required, compared to undertaking multiple standalone RCTs to understand several treatments. MAMS are different to a factorial design in that MAMS treat each arm independent when estimating treatment effects, whereas factorial designs estimate the treatment effects by combining data, this requires an assumption that effects are additive when treatments are combined [107]. Factorial designs are not adaptive trial such there are no interim analysis built in.

Appendix B

Ethics approvals

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Mrs Lauren Bell
LSHTM

2 February 2021

Dear Mrs Lauren Bell

Study Title: Visualising engagement with an behaviour change app for alcohol reduction

LSHTM Ethics Ref: 22701

Thank you for responding to the Observational Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Other	GCP_GCP Certificate	14/01/2016	v1
Investigator CV	Lauren Bell Resume	30/09/2017	v1
Other	GDPR_2018	10/09/2018	v1
Other	Information Security Awareness - Certificate	16/11/2018	v1
Other	IG cert LB	18/11/2018	v1
Investigator CV	CV_Elizabeth_Williamson_24_04_19	24/04/2019	v1
Investigator CV	CV Olga Perski September 2019	30/09/2019	v1
Investigator CV	GCP_Certificate_LB	30/09/2019	v1
Local Approval	Ethics UCL approval	11/10/2019	v1
Investigator CV	CV_Claire Garnett	31/08/2020	v1
Consent form	ethics infromed consent	28/09/2020	v1
Investigator CV	CV for Nav 2019	28/09/2020	v1
Protocol / Proposal	protocol for ethics v2. docx	28/09/2020	v3
Covering Letter	reply to ethics 18022021	18/01/2021	1.0

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

[redacted]

Professor Jimmy Whitworth
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

Improving health worldwide

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Mrs Lauren Bell
LSHTM

27 November 2019

Dear Lauren,

Study Title: Notifications To Improve Engagement (NOTE) trial: a Micro-Randomised Trial to investigate how notifications influence engagement with the Drink Less app

LSHTM Ethics Ref: 17929

Thank you for responding to the Interventions Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Other	GCP_GCP Certificate EW	14/01/2016	1
Safety Information	Drink Less_ Risk assessment	09/10/2018	1
Investigator CV	CV_Elizabeth_Williamson_24_04_19	24/04/2019	1
Protocol / Proposal	NOTE protocol v0.10	23/09/2019	0.10
Other	GCP_Certificate_LB	30/09/2019	1
Investigator CV	Lauren Bell Resume	30/09/2019	1
Investigator CV	CV for HW 2019	30/09/2019	1
Investigator CV	CV_Claire Garnett	30/09/2019	1
Investigator CV	CV Olga Perski September 2019	30/09/2019	1
Sponsor Letter	Sponsorship	30/09/2019	1
Local Approval	ethics approval UCL	11/10/2019	1
Local Approval	FINAL Amendment_Approval_Request_Form	11/10/2019	1
Covering Letter	Response to LSHTM ethics draft FINAL	15/11/2019	1
Covering Letter	Privacy Notice FINAL	15/11/2019	1
Covering Letter	Information sheet for a research project FINAL	15/11/2019	1
Information Sheet	Information sheet for a research project FINAL	15/11/2019	2
Information Sheet	Privacy Notice FINAL	15/11/2019	2

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review

by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

[redacted]

Professor Jimmy Whitworth
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

Improving health worldwide

Bibliography

[1] Charles Abraham et al. "The UK National Institute for Health and Clinical Excellence public health guidance on behaviour change: a brief introduction". In: *Psychology, health & medicine* 14.1 (2009), pp. 1–8.

[2] Niclas Abrahamsson et al. "Video-based CBT-E improves eating patterns in obese patients with eating disorder: A single case multiple baseline study". In: *Journal of behavior therapy and experimental psychiatry* 61 (2018), pp. 104–112.

[3] Urs-Vito Albrecht and Ute von Jan. "Safe, sound and desirable: development of mHealth apps under the stress of rapid life cycles". In: *Mhealth* 3 (2017).

[4] Ghadah Alkhaldi et al. "The effectiveness of prompts to promote engagement with digital interventions: a systematic review". In: *Journal of medical Internet research* 18.1 (2016), e4790.

[5] John P Allen et al. “A review of research on the Alcohol Use Disorders Identification Test (AUDIT)”. In: *Alcoholism: clinical and experimental research* 21.4 (1997), pp. 613–619.

[6] Hannah Brown Amoakoh et al. “How and why front-line health workers (did not) use a multifaceted mHealth intervention to support maternal and neonatal healthcare decision-making in Ghana”. In: *BMJ global health* 4.2 (2019), e001153.

[7] TD Aungst et al. “How to identify, assess and utilise mobile medical applications in clinical practice”. In: *International journal of clinical practice* 68.2 (2014), pp. 155–162.

[8] Vincenzo Bagnardi et al. “Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis”. In: *British journal of cancer* 112.3 (2015), pp. 580–593.

[9] Andreas Balaskas et al. “Ecological momentary interventions for mental health: A scoping review”. In: *PloS one* 16.3 (2021), e0248152.

[10] H Banks et al. “Dynamic modeling of behavior change”. In: *Quarterly of Applied Mathematics* 72.2 (2014), pp. 209–251.

[11] Reuben M Baron and David A Kenny. “The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.” In: *Journal of personality and social psychology* 51.6 (1986), p. 1173.

[12] Colin Barr et al. “The PREEMPT study-evaluating smartphone-assisted n-of-1 trials in patients with chronic pain: study protocol for a randomized controlled trial”. In: *Trials* 16.1 (2015), pp. 1–11.

[13] David W Bates, Adam Landman, and David M Levine. “Health apps and health policy: what is needed?” In: *Jama* 320.19 (2018), pp. 1975–1976.

[14] Ellen Beckjord and Saul Shiffman. “Background for real-time monitoring and intervention related to alcohol use”. In: *Alcohol research: current reviews* 36.1 (2014), p. 9.

[15] Robert Bedbrook and Zachary Byfield. *Digital therapeutics in the primary healthcare setting*. 2018.

[16] Christopher G Beevers et al. “Effectiveness of an internet intervention (Deprexis) for depression in a united states adult sample: A parallel-group pragmatic randomized controlled trial.” In: *Journal of consulting and clinical psychology* 85.4 (2017), p. 367.

[17] Imogen H Bell et al. “Smartphone-based ecological momentary assessment and intervention in a blended coping-focused therapy for distressing voices: Development and case illustration”. In: *Internet interventions* 14 (2018), pp. 18–25.

[18] Lauren Bell et al. “Engagement With a Behavior Change App for Alcohol Reduction: Data Visualization for Longitudinal Observational Study”. In: *Journal of medical Internet research* 22.12 (2020), e23369.

[19] Lauren Bell et al. “Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial”. In: *JMIR research protocols* 9.8 (2020), e18690.

[20] Marvin E Belzer et al. “Adaptive antiretroviral therapy adherence interventions for youth living with HIV through text message and cell phone support with and without incentives: protocol for a sequential multiple assignment randomized trial (SMART)”. In: *JMIR research protocols* 7.12 (2018), e11183.

[21] Marcus Bendtsen et al. “Text messaging interventions for reducing alcohol consumption among risky drinkers: systematic review and meta-analysis”. In: *Addiction* 116.5 (2021), pp. 1021–1033.

[22] Thomas Berger et al. “Evaluating an e-mental health program (“deprexis”) as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial”. In: *Journal of affective disorders* 227 (2018), pp. 455–462.

[23] Thomas Berger et al. “Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help”. In: *Cognitive behaviour therapy* 40.4 (2011), pp. 251–266.

[24] N Bidargaddi et al. “Designing m-Health interventions for precision mental health support”. In: *Translational psychiatry* 10.1 (2020), pp. 1–8.

[25] Niranjan Bidargaddi et al. “To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal

engagement with a mobile health app". In: *JMIR mHealth and uHealth* 6.11 (2018), e10123.

[26] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.

[27] Hans-Hermann Bock. "Clustering methods: a history of k-means algorithms". In: *Selected contributions in data analysis and classification* (2007), pp. 161–172.

[28] Audrey Boruvka et al. "Assessing time-varying causal effect moderation in mobile health". In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1112–1121.

[29] Maged N Kamel Boulos et al. "How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX". In: *Biomedical engineering online* 10.1 (2011), pp. 1–14.

[30] Erin E Brannon et al. "Goal feedback from whom? A physical activity intervention using an N-of-1 RCT". In: *Psychology & health* 33.6 (2018), pp. 701–712.

[31] Babette A Brumback. *Fundamentals of Causal Inference with R: With R*. Chapman and Hall/CRC, 2021.

[32] Lara Bücker et al. "Effects of a depression-focused internet intervention in slot machine gamblers: a randomized controlled trial". In: *PloS one* 13.6 (2018), e0198859.

[33] Alan J Budney et al. “Workshop on the development and evaluation of digital therapeutics for health behavior change: Science, methods, and projects”. In: *JMIR Mental Health* 7.2 (2020), e16751.

[34] Matthew P Buman et al. “BeWell24: development and process evaluation of a smartphone “app” to improve sleep, sedentary, and active behaviors in US Veterans with increased metabolic risk”. In: *Translational Behavioral Medicine* 6.3 (2016), pp. 438–448.

[35] RJ Burns, SS Deschenes, and N Schmitz. “Cyclical relationship between depressive symptoms and diabetes distress in people with Type 2 diabetes mellitus: results from the Montreal Evaluation of Diabetes Treatment Cohort Study”. In: *Diabetic Medicine* 32.10 (2015), pp. 1272–1278.

[36] Robyn Burton et al. “A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective”. In: *The Lancet* 389.10078 (2017), pp. 1558–1580.

[37] Oyungerel Byambasuren et al. “Prescribable mHealth apps identified from an overview of systematic reviews”. In: *NPJ digital medicine* 1.1 (2018), pp. 1–12.

[38] Joseph A Cafazzo et al. “Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study”. In: *Journal of medical Internet research* 14.3 (2012), e2058.

[39] Aimee NC Campbell et al. “Internet-delivered treatment for substance abuse: a multisite randomized controlled trial”. In: *American Journal of Psychiatry* 171.6 (2014), pp. 683–690.

[40] Michelle Campbell et al. “Framework for design and evaluation of complex interventions to improve health”. In: *Bmj* 321.7262 (2000), pp. 694–696.

[41] James Cane, Denise O’Connor, and Susan Michie. “Validation of the theoretical domains framework for use in behaviour change and implementation research”. In: *Implementation science* 7.1 (2012), pp. 1–17.

[42] Stephanie M Carpenter et al. “Developments in mobile health just-in-time adaptive interventions for addiction science”. In: *Current addiction reports* 7.3 (2020), pp. 280–290.

[43] Nicholas Carr. *The shallows: How the internet is changing the way we think, read and remember*. Atlantic Books Ltd, 2010.

[44] Bibhas Chakraborty. “Dynamic treatment regimes for managing chronic health conditions: a statistical perspective”. In: *American journal of public health* 101.1 (2011), pp. 40–45.

[45] Bibhas Chakraborty and EE Moodie. “Statistical methods for dynamic treatment regimes”. In: *Springer-Verlag. doi* 10 (2013), pp. 978–1.

[46] Cynthia A Challener. “FDA framework spurs advanced therapies”. In: *BioPharm International* 1 (2018), pp. 36–38.

[47] Siddharth Chandra, Deborah Scharf, and Saul Shiffman. "Within-day temporal patterns of smoking, withdrawal symptoms, and craving". In: *Drug and alcohol dependence* 117.2-3 (2011), pp. 118–125.

[48] M Chatwin et al. "Randomised crossover trial of telemonitoring in chronic respiratory patients (TeleCRAFT trial)". In: *Thorax* 71.4 (2016), pp. 305–311.

[49] Ioannis A Chatzipavlou, Sofia A Christoforidou, and Maro Vlachopoulou. "A recommended guideline for the development of mHealth Apps". In: *Mhealth* 2 (2016).

[50] Chul-Hyun Cho and Heon-Jeong Lee. "Could digital therapeutics be a game changer in psychiatry?" In: *Psychiatry investigation* 16.2 (2019), p. 97.

[51] Amy M Cohn et al. "Promoting behavior change from alcohol use through mobile technology: the future of ecological momentary assessment". In: *Alcoholism: Clinical and Experimental Research* 35.12 (2011), pp. 2209–2215.

[52] Stephanie Colbert, Louise Thornton, and Robyn Richmond. "Smartphone apps for managing alcohol consumption: a literature review". In: *Addiction Science & Clinical Practice* 15.1 (2020), pp. 1–16.

[53] Linda M Collins and Kari C Kugler. "Optimization of behavioral, biobehavioral, and biomedical interventions". In: *Cham: Springer International Publishing* 10.1007 (2018), pp. 978–973.

[54] Linda M Collins, Susan A Murphy, and Victor Strecher. "The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions". In: *American journal of preventive medicine* 32.5 (2007), S112–S118.

[55] Linda M Collins, Inbal Nahum-Shani, and Daniel Almirall. "Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART)". In: *Clinical Trials* 11.4 (2014), pp. 426–434.

[56] Linda M Collins et al. "A strategy for optimizing and evaluating behavioral interventions". In: *Annals of Behavioral Medicine* 30.1 (2005), pp. 65–73.

[57] Peter Craig et al. "Developing and evaluating complex interventions: the new Medical Research Council guidance". In: *Bmj* 337 (2008).

[58] David Crane et al. "A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial". In: *Scientific reports* 8.1 (2018), pp. 1–11.

[59] David Crane et al. "Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies". In: *Frontiers in public health* 5 (2017), p. 39.

[60] Kristina Elizabeth Curtis, Sudakshina Lahiri, and Katherine Elizabeth Brown. "Targeting parents for childhood weight management:

development of a theory-driven and user-centered healthy eating app". In: *JMIR mHealth and uHealth* 3.2 (2015), e3857.

[61] Jesse Dallery, Rachel N Cassidy, Bethany R Raiff, et al. "Single-case experimental designs to evaluate novel technology-based health interventions". In: *Journal of medical Internet research* 15.2 (2013), e2227.

[62] Jesse Dallery and Bethany R Raiff. "Optimizing behavioral health interventions with single-case designs: from development to dissemination". In: *Translational behavioral medicine* 4.3 (2014), pp. 290–303.

[63] Nina Deliu. "Reinforcement Learning in Modern Biostatistics: Benefits, Challenges and New Proposals". In: (2021).

[64] Walter Dempsey et al. "The stratified micro-randomized trial design: sample size considerations for testing nested causal effects of time-varying treatments". In: *The annals of applied statistics* 14.2 (2020), p. 661.

[65] Laura Dennison et al. "Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study". In: *Journal of medical Internet research* 15.4 (2013), e2583.

[66] Annette DeVito Dabbs et al. "Clinical trials of health information technology interventions intended for patient use: unique issues and considerations". In: *Clinical trials* 10.6 (2013), pp. 896–906.

[67] Peter Diggle et al. *Analysis of longitudinal data*. Oxford university press, 2002.

[68] Jessilyn Dunn, Ryan Runge, and Michael Snyder. “Wearables and the medical revolution”. In: *Personalized medicine* 15.5 (2018), pp. 429–448.

[69] Claire Easthall and Nina Barnett. “Using theory to explore the determinants of medication adherence; moving away from a one-size-fits-all approach”. In: *Pharmacy* 5.3 (2017), p. 50.

[70] Elizabeth Ann Edwards et al. “Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps”. In: *BMJ open* 6.10 (2016), e012447.

[71] Garry Egger and John Dixon. “Should obesity be the main game? Or do we need an environmental makeover to combat the inflammatory and chronic disease epidemics?” In: *obesity reviews* 10.2 (2009), pp. 237–249.

[72] Anna Essén et al. “Health app policy: international comparison of nine countries’ approaches”. In: *NPJ digital medicine* 5.1 (2022), pp. 1–10.

[73] Gunther Eysenbach, Consort-EHEALTH Group, et al. “CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions”. In: *Journal of medical Internet research* 13.4 (2011), e1923.

[74] US FDA. *Policy for device software functions and mobile medical applications: guidance for industry and Food and Drug Administration staff*. 2019.

[75] Joseph Firth, John Torous, et al. “Smartphone apps for schizophrenia: a systematic review”. In: *JMIR mHealth and uHealth* 3.4 (2015), e4930.

[76] Anja Fischer et al. “An online programme to reduce depression in patients with multiple sclerosis: a randomised controlled trial”. In: *The Lancet Psychiatry* 2.3 (2015), pp. 217–223.

[77] Martin Fitzgerald and Tracy McClelland. “What makes a mobile app successful in supporting health behaviour change?” In: *Health education journal* 76.3 (2017), pp. 373–381.

[78] Garrett Fitzmaurice et al. *Longitudinal data analysis*. CRC press, 2008.

[79] Theresa Fleming et al. “Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety”. In: *Journal of medical Internet research* 20.6 (2018), e9275.

[80] Gordon Forbes et al. “MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial”. In: *BMJ open* 10.3 (2020), e030164.

[81] Mohammad H Forouzanfar et al. “Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic

analysis for the Global Burden of Disease Study 2015". In: *The lancet* 388.10053 (2016), pp. 1659–1724.

[82] Lauren A Fowler, Sidney L Holt, and Deepti Joshi. "Mobile technology-based interventions for adult users of alcohol: a systematic review of the literature". In: *Addictive Behaviors* 62 (2016), pp. 25–34.

[83] Emily Anne Fulton et al. "StopApp: using the behaviour change wheel to develop an app to increase uptake and attendance at NHS Stop Smoking Services". In: *Healthcare*. Vol. 4. 2. MDPI. 2016, p. 31.

[84] Claire Garnett et al. "Evaluating the effectiveness of the smartphone app, Drink Less, compared with the NHS alcohol advice webpage, for the reduction of alcohol consumption among hazardous and harmful adult drinkers in the UK at 6-month follow-up: protocol for a randomised controlled trial". In: *Addiction* 116.2 (2021), pp. 412–425.

[85] Claire Garnett et al. "Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method". In: *JMIR mHealth and uHealth* 3.2 (2015), e3895.

[86] Claire Garnett et al. "Predictors of engagement, response to follow up, and extent of alcohol reduction in users of a smartphone app (drink less): secondary analysis of a factorial randomized controlled trial". In: *JMIR mHealth and uHealth* 6.12 (2018), e11175.

[87] Claire Garnett et al. “Refining the content and design of an alcohol reduction app, Drink Less, to improve its usability and effectiveness: a mixed methods approach”. In: *F1000Research* 10 (2021).

[88] Claire Garnett et al. “The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers”. In: *Translational behavioral medicine* 9.2 (2019), pp. 296–307.

[89] Soumi Ghosh and Sanjay Kumar Dubey. “Comparative analysis of k-means and fuzzy c-means algorithms”. In: *International Journal of Advanced Computer Science and Applications* 4.4 (2013).

[90] Sander Greenland. “Interpretation and choice of effect measures in epidemiologic analyses”. In: *American journal of epidemiology* 125.5 (1987), pp. 761–768.

[91] Nicola Grissom and Seema Bhatnagar. “Habituation to repeated stress: get used to it”. In: *Neurobiology of learning and memory* 92.2 (2009), pp. 215–224.

[92] Garrett Grolemund and Hadley Wickham. “Dates and times made easy with lubridate”. In: *Journal of statistical software* 40 (2011), pp. 1–25.

[93] Quinn Grundy. “A Review of the Quality and Impact of Mobile Health Apps”. In: *Annual Review of Public Health* 43 (2022), pp. 117–134.

[94] Kate Guastaferro, Jillian C Strayhorn, and Linda M Collins. “The multiphase optimization strategy (MOST) in child maltreatment pre-

vention research". In: *Journal of child and family studies* 30.10 (2021), pp. 2481–2491.

[95] David H Gustafson et al. "A smartphone application to support recovery from alcoholism: a randomized clinical trial". In: *JAMA psychiatry* 71.5 (2014), pp. 566–572.

[96] Eduardo Hariton and Joseph J Locascio. "Randomised controlled trials—the gold standard for effectiveness research". In: *BJOG: an international journal of obstetrics and gynaecology* 125.13 (2018), p. 1716.

[97] Eric B Hekler et al. "Advancing models and theories for digital behavior change interventions". In: *American journal of preventive medicine* 51.5 (2016), pp. 825–832.

[98] Miguel A Hernán and James M Robins. *Causal inference*. 2010.

[99] Kristin E Heron and Joshua M Smyth. "Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments". In: *British journal of health psychology* 15.1 (2010), pp. 1–39.

[100] Emily Herrett et al. "Study protocol for statin web-based investigation of side effects (StatinWISE): a series of randomised controlled N-of-1 trials comparing atorvastatin and placebo in UK primary care". In: *BMJ open* 7.12 (2017), e016604.

[101] Zhexue Huang. "Extensions to the k-means algorithm for clustering large data sets with categorical values". In: *Data mining and knowledge discovery* 2.3 (1998), pp. 283–304.

[102] Zhexue Huang and Michael K Ng. "A fuzzy k-modes algorithm for clustering categorical data". In: *IEEE transactions on Fuzzy Systems* 7.4 (1999), pp. 446–452.

[103] Zhilian Huang et al. "Clinical relevance of smartphone apps for diabetes management: A global overview". In: *Diabetes/metabolism research and reviews* 34.4 (2018), e2990.

[104] Casper D Hulshof and Ton De Jong. "Using just-in-time information to support scientific discovery learning in a computer-based simulation". In: *Interactive Learning Environments* 14.1 (2006), pp. 79–94.

[105] George M Hunt and Nathan H Azrin. "A community-reinforcement approach to alcoholism". In: *Behaviour research and therapy* 11.1 (1973), pp. 91–104.

[106] John PA Ioannidis. "Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials". In: *Jama* 279.4 (1998), pp. 281–286.

[107] Thomas Jaki and Despina Vasileiou. "Factorial versus multi-arm multistage designs for clinical trials with multiple treatments". In: *Statistics in Medicine* 36.4 (2017), pp. 563–580.

[108] Gareth James et al. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[109] Eunjoo Jeon and Hyeoun-Ae Park. "Factors affecting acceptance of smartphone application for management of obesity". In: *Healthcare informatics research* 21.2 (2015), pp. 74–82.

[110] Eileen FS Kaner et al. “Effectiveness of brief alcohol interventions in primary care populations”. In: *Cochrane database of systematic reviews* 2 (2018).

[111] Eileen FS Kaner et al. “Intervention for excessive alcohol consumption in primary health care: attitudes and practices of English general practitioners”. In: *Alcohol and Alcoholism* 34.4 (1999), pp. 559–566.

[112] Alboukadel Kassambara et al. “survminer: Drawing Survival Curves using ‘ggplot2’”. In: *R package version 0.3* 1 (2017).

[113] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. “On the complexity of A/B testing”. In: *Conference on Learning Theory*. PMLR. 2014, pp. 461–481.

[114] Ruth H Keogh et al. “Analysis of longitudinal studies with repeated outcome measures: adjusting for time-dependent confounding using conventional methods”. In: *American journal of epidemiology* 187.5 (2018), pp. 1085–1092.

[115] Ariane Kerst, Jürgen Zielasek, and Wolfgang Gaebel. “Smartphone applications for depression: a systematic literature review and a survey of health care professionals’ attitudes towards their use in clinical practice”. In: *European archives of psychiatry and clinical neuroscience* 270.2 (2020), pp. 139–152.

[116] Predrag Klasnja et al. “Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps”. In: *Annals of Behavioral Medicine* 53.6 (2019), pp. 573–582.

[117] Predrag Klasnja et al. “Microrandomized trials: An experimental design for developing just-in-time adaptive interventions.” In: *Health Psychology* 34.S (2015), p. 1220.

[118] Jan Philipp Klein et al. “The EVIDENT-trial: protocol and rationale of a multicenter randomized controlled trial testing the effectiveness of an online-based psychological intervention”. In: *BMC psychiatry* 13.1 (2013), pp. 1–10.

[119] Jan-Niklas Kramer et al. “Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial”. In: *JMIR research protocols* 8.1 (2019), e11540.

[120] Tobias Krieger et al. “Evaluating an e-mental health program (“deprexis”) as adjunctive treatment tool in psychotherapy for depression: design of a pragmatic randomized controlled trial”. In: *BMC psychiatry* 14.1 (2014), pp. 1–8.

[121] Kari C Kugler et al. “Application of the multiphase optimization strategy to a pilot study: an empirical example targeting obesity among children of low-income mothers”. In: *BMC Public Health* 16.1 (2016), pp. 1–10.

[122] Santosh Kumar et al. “Mobile health technology evaluation: the mHealth evidence workshop”. In: *American journal of preventive medicine* 45.2 (2013), pp. 228–236.

[123] Sarah Lagan, Lev Sandler, and John Torous. “Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps”. In: *BMJ open* 11.3 (2021), e047001.

[124] Rashmi Lakshminarayana et al. “Smartphone-and internet-assisted self-management and adherence tools to manage Parkinson’s disease (SMART-PD): study protocol for a randomised controlled trial (v7; 15 August 2014)”. In: *Trials* 15.1 (2014), pp. 1–11.

[125] Sylvie D Lambert et al. “Adaptive web-based stress management programs among adults with a cardiovascular disease: A pilot Sequential Multiple Assignment Randomized Trial (SMART)”. In: *Patient Education and Counseling* 105.6 (2022), pp. 1587–1597.

[126] The Lancet. *Is digital medicine different?* 2018.

[127] Philip W Lavori and Ree Dawson. “Dynamic treatment regimes: practical design considerations”. In: *Clinical trials* 1.1 (2004), pp. 9–20.

[128] Lisa M Law and James MS Wason. “Design of telehealth trials–introducing adaptive approaches”. In: *International Journal of Medical Informatics* 83.12 (2014), pp. 870–880.

[129] Theodore T Lee and Aaron S Kesselheim. “US Food and Drug Administration precertification pilot program for digital health software: weighing the benefits and risks”. In: *Annals of internal medicine* 168.10 (2018), pp. 730–732.

[130] Valentina Lenarduzzi and Davide Taibi. “MVP explained: A systematic mapping study on the definitions of minimal viable product”. In:

2016 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE. 2016, pp. 112–119.

[131] Thomas Lorchan Lewis and Jeremy C Wyatt. "mHealth and mobile medical apps: a framework to assess risk and promote safer use". In: *Journal of medical Internet research* 16.9 (2014), e3133.

[132] Peng Liao et al. "Just-in-time but not too much: Determining treatment timing in mobile health". In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2.4 (2018), pp. 1–21.

[133] Peng Liao et al. "Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.1 (2020), pp. 1–22.

[134] Elizabeth O Lillie et al. "The n-of-1 clinical trial: the ultimate strategy for individualizing medicine?" In: *Personalized medicine* 8.2 (2011), pp. 161–173.

[135] Roderick J Little and Roger J Lewis. "Estimands, estimators, and estimates". In: *JAMA* 326.10 (2021), pp. 967–968.

[136] Pernille Lunde et al. "The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses". In: *Journal of medical Internet research* 20.5 (2018), e9751.

[137] J MacQueen. "Classification and analysis of multivariate observations". In: *5th Berkeley Symp. Math. Statist. Probability*. 1967, pp. 281–297.

[138] Jamie M Marshall, Debra A Dunstan, and Warren Bartik. "The digital psychiatrist: in search of evidence-based apps for anxiety and depression". In: *Frontiers in Psychiatry* 10 (2019), p. 831.

[139] Claire McCallum, John Rooksby, Cindy M Gray, et al. "Evaluating the impact of physical activity apps and wearables: interdisciplinary review". In: *JMIR mHealth and uHealth* 6.3 (2018), e9054.

[140] Margaret McCartney. "Margaret McCartney: Health technology and the modern inverse care law". In: *BMJ* 362 (2018).

[141] Jennifer B McClure et al. "The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial". In: *Journal of medical Internet research* 15.3 (2013), e2508.

[142] Meghan E McDevitt-Murphy, Matthew T Luciano, and Rebecca J Zakarian. "Use of ecological momentary assessment and intervention in treatment with adults". In: *Focus* 16.4 (2018), pp. 370–375.

[143] Fiona H McKay et al. "Evaluating mobile phone applications for health behaviour change: A systematic review". In: *Journal of telemedicine and telecare* 24.1 (2018), pp. 22–30.

[144] Geoffrey J McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 2007.

[145] Xiao-Li Meng. “Discussion: Should a Working Model Actually” Work”?” In: *International Statistical Review/Revue Internationale de Statistique* 84.3 (2016), pp. 362–367.

[146] Steven E Meredith, Sheila M Alessi, and Nancy M Petry. “Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review”. In: *Advanced health care technologies* 1 (2015), p. 47.

[147] Nadine Messerli-Bürgy, Jürgen Barth, and Thomas Berger. “The InterHerz project-a web-based psychological treatment for cardiac patients with depression: study protocol of a randomized controlled trial”. In: *Trials* 13.1 (2012), pp. 1–7.

[148] Björn Meyer et al. “Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial”. In: *Journal of medical Internet research* 11.2 (2009), e1151.

[149] Susan Michie, Maartje M Van Stralen, and Robert West. “The behaviour change wheel: a new method for characterising and designing behaviour change interventions”. In: *Implementation science* 6.1 (2011), pp. 1–12.

[150] Susan Michie et al. “Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop”. In: *Journal of medical Internet research* 19.6 (2017), e7126.

[151] Kathryn R Middleton, Stephen D Anton, and Michal G Perri. "Long-term adherence to health behavior change". In: *American journal of lifestyle medicine* 7.6 (2013), pp. 395–404.

[152] Michael L Millenson et al. "Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis". In: *Diagnosis* 5.3 (2018), pp. 95–105.

[153] William R Miller, Robert J Meyers, et al. "The community-reinforcement approach". In: *Psychosocial Treatments* (2004), pp. 59–70.

[154] Fabian Moerchen, Patrick Ernst, and Giovanni Zappella. "Personalizing natural language understanding using multi-armed bandits and implicit feedback". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020, pp. 2661–2668.

[155] Steffen Moritz et al. "A randomized controlled trial of internet-based therapy in depression". In: *Behaviour research and therapy* 50.7-8 (2012), pp. 513–521.

[156] Tim P Morris, Ian R White, and Michael J Crowther. "Using simulation studies to evaluate statistical methods". In: *Statistics in medicine* 38.11 (2019), pp. 2074–2102.

[157] Cecily Morrison, Gavin Doherty, et al. "Analyzing engagement in a web-based intervention platform through visualizing log-data". In: *Journal of medical Internet research* 16.11 (2014), e3575.

[158] Elizabeth Murray et al. “Methodological challenges in online trials”. In: *Journal of Medical Internet Research* 11.2 (2009), e1052.

[159] Inbal Nahum-Shani, John J Dziak, and David W Wetter. “MCMTC: A Pragmatic Framework for Selecting an Experimental Design to Inform the Development of Digital Interventions”. In: *Frontiers in Digital Health* 4 (2022).

[160] Inbal Nahum-Shani et al. “Engagement in digital interventions.” In: *American Psychologist* (2022).

[161] Inbal Nahum-Shani et al. “Hybrid Experimental Designs for Intervention Development: What, Why, and How”. In: *Advances in Methods and Practices in Psychological Science* 5.3 (2022), p. 25152459221114279.

[162] Inbal Nahum-Shani et al. “Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support”. In: *Annals of Behavioral Medicine* 52.6 (2018), pp. 446–462.

[163] Timothy NeCamp et al. “Assessing real-time moderation for developing adaptive mobile health interventions for medical interns: micro-randomized trial”. In: *Journal of medical Internet research* 22.3 (2020), e15033.

[164] PH NICE. “Behaviour change: Individual approaches”. In: *London: National Institute for Health and Care Excellence* (2014).

[165] Conor O'Brien et al. "Should I send this notification? Optimizing push notifications decision making by modeling the future". In: *arXiv preprint arXiv:2202.08812* (2022).

[166] Reiko Onodera and Shintaro Sengoku. "Innovation process of mHealth: An overview of FDA-approved mobile medical applications". In: *International journal of medical informatics* 118 (2018), pp. 65–71.

[167] World Health Organization et al. "mHealth: new horizons for health through mobile technologies." In: *mHealth: new horizons for health through mobile technologies.* (2011).

[168] Philip Pallmann et al. "Adaptive designs in clinical trials: why use them, and how to run and report them". In: *BMC medicine* 16.1 (2018), pp. 1–15.

[169] Sungmee Park and Sundaresan Jayaraman. "Wearables: Fundamentals, advancements, and a roadmap for the future". In: *Wearable sensors*. Elsevier, 2021, pp. 3–27.

[170] Christine A Pellegrini et al. "Optimization of remotely delivered intensive lifestyle treatment for obesity using the Multiphase Optimization Strategy: Opt-IN study protocol". In: *Contemporary clinical trials* 38.2 (2014), pp. 251–259.

[171] Olga Perski et al. "Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis". In: *Translational behavioral medicine* 7.2 (2017), pp. 254–267.

[172] Olga Perski et al. “Do daily fluctuations in psychological and app-related variables predict engagement with an alcohol reduction app? A series of N-Of-1 studies”. In: *JMIR mHealth and uHealth* 7.10 (2019), e14098.

[173] Maya L Petersen and Mark J van der Laan. “Causal models and learning from data: integrating causal modeling and statistical estimation”. In: *Epidemiology (Cambridge, Mass.)* 25.3 (2014), p. 418.

[174] Quynh Pham, David Wiljer, Joseph A Cafazzo, et al. “Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods”. In: *JMIR mHealth and uHealth* 4.3 (2016), e5720.

[175] Quynh Pham et al. “A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review”. In: *JMIR mHealth and uHealth* 7.1 (2019), e11941.

[176] Nada Y Philip et al. “Internet of Things for in-home health monitoring systems: current advances, challenges and future directions”. In: *IEEE Journal on Selected Areas in Communications* 39.2 (2021), pp. 300–310.

[177] Siobhan M Phillips et al. “Optimization of a technology-supported physical activity intervention for breast cancer survivors: Fit2Thrive study protocol”. In: *Contemporary clinical trials* 66 (2018), pp. 9–19.

[178] Daniel Philpott, Aziz Guergachi, and Karim Keshavjee. “Design and validation of a platform to evaluate mHealth apps”. In: *Informatics*

for Health: Connected Citizen-Led Wellness and Population Health. IOS Press, 2017, pp. 3–7.

[179] Martin Potthast et al. "Crowdsourcing a large corpus of clickbait on twitter". In: *Proceedings of the 27th international conference on computational linguistics*. 2018, pp. 1498–1507.

[180] Tianchen Qian et al. "Estimating time-varying causal excursion effects in mobile health with binary outcomes". In: *Biometrika* 108.3 (2021), pp. 507–527.

[181] Tianchen Qian et al. "The microrandomized trial for developing digital interventions: Experimental design and data analysis considerations." In: *Psychological Methods* (2022).

[182] Francis Quinn, Marie Johnston, and Derek W Johnston. "Testing an integrated behavioural and biomedical model of disability in N-of-1 studies with chronic pain". In: *Psychology & Health* 28.12 (2013), pp. 1391–1406.

[183] Mashfiqui Rabbi et al. "Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults". In: *JMIR mHealth and uHealth* 3.2 (2015), e4160.

[184] Mashfiqui Rabbi et al. "SARA: a mobile app to engage users in health data collection". In: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings*

of the 2017 ACM International Symposium on Wearable Computers. 2017, pp. 781–789.

[185] Mashfiqui Rabbi et al. "Toward increasing engagement in substance use data collection: development of the Substance Abuse Research Assistant app and protocol for a microrandomized trial using adolescents and emerging adults". In: *JMIR research protocols* 7.7 (2018), e9850.

[186] Catharine H Rankin et al. "Habituation revisited: an updated and revised description of the behavioral characteristics of habituation". In: *Neurobiology of learning and memory* 92.2 (2009), pp. 135–138.

[187] FDA NEWS RELEASE. *FDA permits marketing of mobile medical application for substance use disorder*. 2017. URL: https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-mobile-medical-application-substance-use-disorder (visited on 06/04/2022).

[188] Eric Ries. "Minimum viable product: a guide". In: *Startup lessons learned* 3 (2009), p. 1.

[189] William T Riley et al. "Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise". In: *Clinical and translational medicine* 2.1 (2013), pp. 1–6.

[190] Anna L Roberts et al. "Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis". In: *Journal of Cancer Survivorship* 11.6 (2017), pp. 704–719.

[191] James M Robins, Miguel Angel Hernan, and Babette Brumback. *Marginal structural models and causal inference in epidemiology*. 2000.

[192] Mary AM Rogers et al. "Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability". In: *Journal of medical Internet research* 19.3 (2017), e7111.

[193] Donald B Rubin. "Estimating causal effects of treatments in randomized and nonrandomized studies." In: *Journal of educational Psychology* 66.5 (1974), p. 688.

[194] Jason M Satterfield. *Thinking Outside the Visit: Digitally Extending the Reach of Behavioral Health*. 2016.

[195] John B Saunders et al. "Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II". In: *Addiction* 88.6 (1993), pp. 791–804.

[196] Alexander O Savi et al. "Delaying access to a problem-skipping option increases effortful practice: Application of an A/B test in large-scale online learning". In: *Computers & Education* 119 (2018), pp. 84–94.

[197] Kenneth F Schulz, Douglas G Altman, and David Moher. "CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials". In: *Trials* 11.1 (2010), pp. 1–8.

[198] Steven L Scott. “A modern Bayesian look at the multi-armed bandit”. In: *Applied Stochastic Models in Business and Industry* 26.6 (2010), pp. 639–658.

[199] Stephen Senn. “Seven myths of randomisation in clinical trials”. In: *Statistics in medicine* 32.9 (2013), pp. 1439–1450.

[200] Saul Shiffman. “Ecological momentary assessment (EMA) in studies of substance use.” In: *Psychological assessment* 21.4 (2009), p. 486.

[201] Saul Shiffman, Arthur A Stone, and Michael R Hufford. “Ecological momentary assessment”. In: *Annu. Rev. Clin. Psychol.* 4 (2008), pp. 1–32.

[202] Bonnie Sibbald and Chris Roberts. “Understanding controlled trials crossover trials”. In: *Bmj* 316.7146 (1998), pp. 1719–1720.

[203] Daniel D Sjoberg et al. “Reproducible Summary Tables with the gtsummary Package.” In: *R Journal* 13.1 (2021).

[204] Kathryn Skivington et al. “A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance”. In: *bmj* 374 (2021).

[205] Nadia Solaro et al. “GenForImp: The Forward Imputation: A Sequential Distance-Based Approach for Imputing Missing Data”. In: (2015).

[206] Bonnie Spring et al. “A factorial experiment to optimize remotely delivered behavioral treatment for obesity: results of the Opt-IN study”. In: *Obesity* 28.9 (2020), pp. 1652–1662.

[207] Katarzyna Stawarz et al. “User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews”. In: *Journal of medical Internet research* 20.6 (2018), e10120.

[208] Arthur A Stone and Saul Shiffman. “Ecological momentary assessment (EMA) in behavorial medicine.” In: *Annals of behavioral medicine* (1994).

[209] Margaret Sullivan Pepe and Garnet L Anderson. “A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data”. In: *Communications in statistics-simulation and computation* 23.4 (1994), pp. 939–951.

[210] Oleksandr Sverdlov et al. “Digital therapeutics: an integral component of digital innovation in drug development”. In: *Clinical Pharmacology & Therapeutics* 104.1 (2018), pp. 72–80.

[211] Matthew R Sydes et al. “Flexible trial design in practice-stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial”. In: *Trials* 13.1 (2012), pp. 1–14.

[212] R Core Team et al. “R: A language and environment for statistical computing”. In: (2013).

[213] Ildiko Tombor et al. “Development of SmokeFree Baby: a smoking cessation smartphone app for pregnant smokers”. In: *Translational behavioral medicine* 6.4 (2016), pp. 533–545.

[214] A Janet Tomiyama. “Weight stigma is stressful. A review of evidence for the Cyclic Obesity/Weight-Based Stigma model”. In: *Appetite* 82 (2014), pp. 8–15.

[215] John Torous and Honor Hsin. “Empowering the digital therapeutic relationship: virtual clinics for digital health interventions”. In: *NPJ digital medicine* 1.1 (2018), pp. 1–3.

[216] Andrea C Tricco et al. “PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation”. In: *Annals of internal medicine* 169.7 (2018), pp. 467–473.

[217] Conal Twomey, Gary O’Reilly, and Björn Meyer. “Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: a meta-analysis”. In: *Psychiatry research* 256 (2017), pp. 371–377.

[218] Harriet Unsworth et al. “The NICE evidence standards framework for digital health and care technologies–developing and maintaining an innovative evidence framework with global impact”. In: *Digital health* 7 (2021), p. 20552076211018617.

[219] Teruhisa Uwatoko et al. “Healthy Campus Trial: a multiphase optimization strategy (MOST) fully factorial trial to optimize the smartphone cognitive behavioral therapy (CBT) app for mental health promotion among university students: study protocol for a randomized controlled trial”. In: *Trials* 19 (2018), pp. 1–16.

[220] Anthony Vance et al. “Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments”. In: *MIS Quarterly* 42.2 (2018), pp. 355–380.

[221] Sofia S Villar, Jack Bowden, and James Wason. “Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges”. In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 30.2 (2015), p. 199.

[222] Ashley Walton et al. “Optimizing digital integrated care via micro-randomized trials”. In: *Clinical Pharmacology & Therapeutics* 104.1 (2018), pp. 53–58.

[223] Liyuan Wang and Lynn Carol Miller. “Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review”. In: *Health Communication* 35.12 (2020), pp. 1531–1544.

[224] James Wason et al. “Some recommendations for multi-arm multi-stage trials”. In: *Statistical methods in medical research* 25.2 (2016), pp. 716–727.

[225] Hadley Wickham. “An introduction to ggplot: An implementation of the grammar of graphics in R”. In: *Statistics* (2006), pp. 1–8.

[226] Hadley Wickham et al. “dplyr: A Grammar of Data Manipulation. R package version 0.7. 6”. In: *Computer software]. https://CRAN. R-project. org/package= dplyr* (2018).

[227] Charlotte Williamson et al. “Smartphone-based alcohol interventions: A systematic review on the role of notifications in changing behaviors toward alcohol”. In: *Substance Abuse* 43.1 (2022), pp. 1231–1244.

[228] Elizabeth Williamson et al. “Propensity scores: from naive enthusiasm to intuitive understanding”. In: *Statistical methods in medical research* 21.3 (2012), pp. 273–293.

[229] Graeme B Wilson et al. “Intervention against excessive alcohol consumption in primary health care: a survey of GPs’ attitudes and practices in England 10 years on”. In: *Alcohol and Alcoholism* 46.5 (2011), pp. 570–577.

[230] Jeffrey C Wong. “Computational causal inference”. In: *arXiv preprint arXiv:2007.10979* (2020).

[231] www.mobihealthnews.com. *Roche’s Accu-Chek diabetes management app receives its fifth FDA recall*. 2018. URL: `https://www.mobihealthnews.com/content/roche%E2%80%99s-accu-chek-diabetes-management-app-receives-its-fifth-fda-recall` (visited on 06/04/2022).

[232] Ya Xu et al. “From infrastructure to culture: A/B testing challenges in large scale social networks”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015, pp. 2227–2236.

[233] Jiayu Yao et al. “Power Constrained Bandits”. In: *Machine Learning for Healthcare Conference*. PMLR. 2021, pp. 209–259.

[234] Lucy Yardley et al. “The person-based approach to enhancing the acceptability and feasibility of interventions”. In: *Pilot and feasibility studies* 1.1 (2015), pp. 1–7.

[235] Sunmoo Yoon et al. “Using behavioral analytics to increase exercise: A randomized N-of-1 study”. In: *American journal of preventive medicine* 54.4 (2018), pp. 559–567.

[236] Achim Zeileis and Gabor Grothendieck. “zoo: S3 infrastructure for regular and irregular time series”. In: *arXiv preprint math/0505527* (2005).

[237] Jing Zhao, Becky Freeman, Mu Li, et al. “Can mobile phone apps influence people’s health behavior change? An evidence review”. In: *Journal of medical Internet research* 18.11 (2016), e5692.

[238] Feiyun Zhu and Peng Liao. “Effective warm start for the online actor-critic reinforcement learning based mhealth intervention”. In: *arXiv preprint arXiv:1704.04866* (2017).

[239] Rüdiger Zwerenz et al. “Enhancing inpatient psychotherapeutic treatment with online self-help: study protocol for a randomized controlled trial”. In: *Trials* 16.1 (2015), pp. 1–7.