

# **The public health potential of mobile applications to increase physical activity**

By Paulina Bondaronek

Thesis submitted for the Degree of Doctor of Philosophy

University College London

Research Department of Primary Care and Population Health

December 2019

## **Declaration**

I, Paulina Bondaronek, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

The following work was carried out at University College London, under the supervision of Professor Elizabeth Murray and Dr Fiona Hamilton (Department of Primary Care Health Sciences).

This thesis has not been submitted, in whole or in part, for any other degree, diploma or qualification at any other University. My work was funded by a studentship from the Medical Research Council.

This thesis does not exceed the limit of 100,000 words specified by the Degree Committee.

Correspondence concerning this thesis should be addressed to Paulina Bondaronek, p.bondaronek@gmail.com

Signed, 12th December 2019

## Acknowledgements

I would like to express my genuine gratitude to my supervisors, Elizabeth Murray and Fiona Hamilton for their guidance, patience and encouragement. The process of conducting this PhD felt like a military training for my brain increasing my critical and analytical thinking fivefold, my resilience sevenfold, and my motivation and aspirations to work in this field tenfold. Thanks to you.

I would like to especially thank April Slee for being such a great hands-on teacher and for opening my mind. Your work ethics as well as your world-view and the ability to widen my horizon is priceless. You are amazing!

My sincere thanks goes to Henry Potts whom I consider my unofficial mentor. Thank you for your support, enthusiasm and stimulating conversations!

Thank you to the members of the Behavioural Insights at PHE. I consider myself incredibly lucky to work within such a stimulating, enjoyable, motivating, and supportive environment.

A special thank you to Marta Buszewicz for her support, encouragement, especially around my first documentary premiere which she attended, giving me a boost of confidence.

Thank you to my fellow PhD students and colleagues who became friends along the way: Hamad Bastaki, Emma Dunphy, Shoba Poduval, Lorraine McDonagh, Katrin Hoffmann, and Marion Lean. Thank you for making my PhD journey mind-opening and full of laughter (through tears sometimes). An amazing bunch you are.

Thank you to Nikki Newhouse and Olga Perski. Talking *digital health research* trials and tribulations contributed substantially to my thinking, and your support contributed significantly to my wellbeing.

Thank you to my lovely friends: Veronika Kaserer for sharing your wisdom, depth and entertainment, Sarah Jane Williams for being not only an attentive ear during tough times but also immense fun, Ada Perkowska and Justyna Dziejcz for stimulating conversations, and Mario Laurenza for being so supportive. Thanks to Kath Leone for

our meetings at the Ladies Pond and the Parliament Hill Lido. Our swims and conversations were so supportive and energising.

I would like to express my gratitude to Lou Atkins and Gill Forbes for providing their expertise unconditionally. It helped me gain more confidence in my research. I appreciate your help!

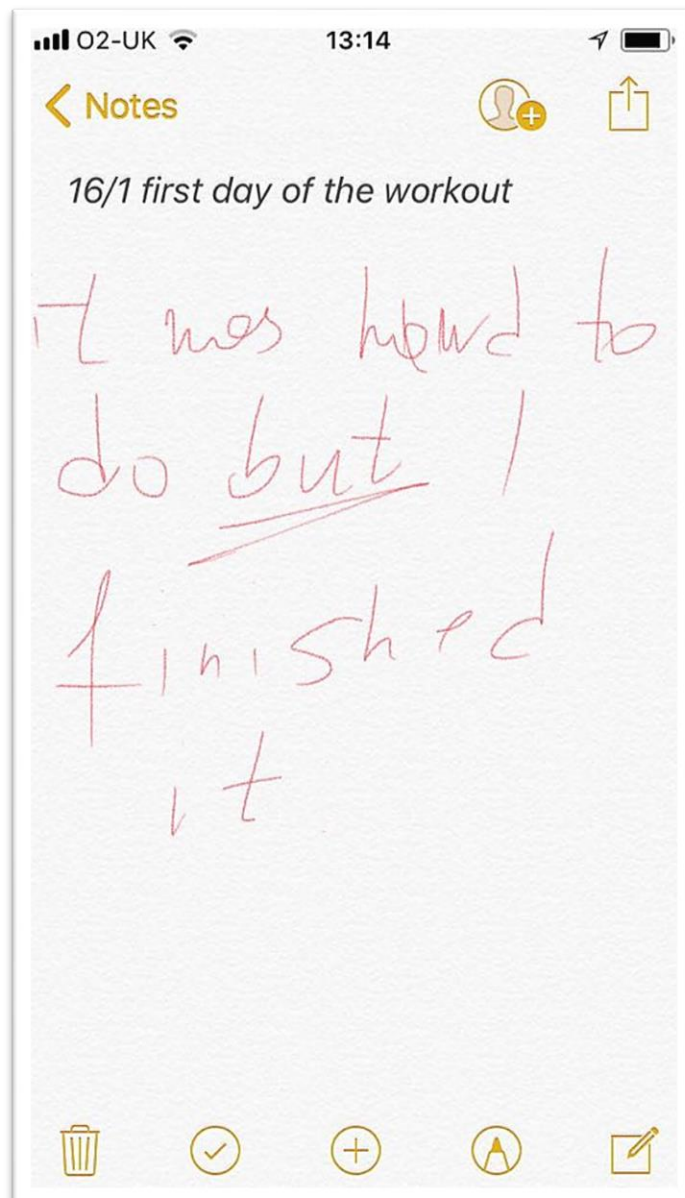
Thank you so much to my Practical Philosophy Group, especially Jonathan Rookcroft, Maria Fernandes, Caroline Ings-Chambers, and Barbara Karban. Discussing wisdom, presence, bliss, truth, and contentment on Saturday mornings created solid foundation for my journey towards what matters.

Thank you to my sister, Karolina Olga Bondaronek, for being the best sister I could ever ask for. I am so grateful for your emotional support and immense empathy. Thank you to my mum, Joanna Bondaronek. You taught me independence and love for knowledge. I would have not been able to embark on this journey if it wasn't for the strength I believe I had. Thanks to you. Thank you to my dad, I think you would be proud.

Thank you to Edna and Tim Anderson for being so welcoming. Your immense support, generosity and enthusiasm helped me so much. You are inspiring.

Finally, I would like to enormously thank my partner Tristan Anderson. Your love, strength, patience and unwavering support provided me with incredible energy, every single day. Thank you for believing in me. You are a truly amazing human being.





*"It was hard to do but I finished it"* - participant 001-CHR, multimedia diary (Study 4). It summarises accurately my self-reflection on the process of finalising this thesis

## **Abstract**

**Background:** Physical activity (PA) is an important behavioural determinant of morbidity and mortality and is a public health priority. The accessibility, convenience and wide reach of mobile applications (apps) makes these digital interventions a potential mode for delivering PA interventions at scale. At the end of 2017 there were 325,000 health apps available publicly, with “fitness” apps being the largest category of all health apps. However, most apps on the market have not been evaluated and little is known about their quality.

**Aim:** This PhD investigated the public health potential of publicly available PA apps.

**Methods:** The following studies were conducted: 1) a review and content analysis of the most popular PA apps on the market to assess their quality, defined as safety, likely efficacy and positive user experience; 2) a study using regression models to determine the association between popularity and quality of those apps; 3) a feasibility crossover trial assessing two apps for increasing PA; and 4) a qualitative study assessing the acceptability of the trial procedures and exploring the experiences of the two PA apps.

**Results:** Popular apps had high usability but there were issues around their safety and likely efficacy. Popularity was not associated with likely efficacy. The feasibility trial and the qualitative study showed that such a trial would be feasible and acceptable to participants. The enablers and barriers to increasing exercise using the apps were identified.

**Conclusion:** The discrepancy between quality and popularity represents a missed opportunity for behaviour change interventions. Hence, the public health impact of PA apps is unlikely to be achieved when market forces “prescribe” what is used by the public. The motivation to use the apps varied substantially and it is important to identify when, for whom, and in what context PA apps are most likely to facilitate behaviour change.

## **Impact statement**

I intend this thesis to be used as a pragmatic guide for those interested in evaluating digital health interventions. Those who find my work useful are welcome to get in touch if they have further questions.

The future impact envisaged is in the areas of applying behavioural science to digital health, and the evaluation of existing digital interventions using novel alternatives to randomised controlled trials.

The three main priorities of the Secretary of State for Health and Social Care are 1) workforce, 2) technology, 3) prevention. Two of these are addressed in this thesis. In relation to prevention, new research evidence has been established showing that even small increases in PA can prolong life and reduce morbidity. In relation to technology, the potential opportunities to support health behaviour change are vast but these are yet to be realised as the digital health market expansion has overtaken the evidence for their effectiveness. Hence, this thesis is important when considering current health politics. This impact statement outlines the areas in which this research is likely to have the clearest impact, and where I was able to use my research to create impact.

First, through Research Council UK Policy Internship Scheme, I first completed a placement in Public Health England (PHE) Behaviour Insights Team which led to my employment by the Team. This gave me an opportunity to promote the findings of the consecutive studies to my Team and also senior colleagues at PHE. Consequently, I applied the knowledge gained and methods learnt to issues relevant to public health. For example, the findings of my research were included in the presentation for the WHO mActive sessions in Geneva this year.

Secondly, I was given the responsibility to develop and test a quality assessment framework for apps included on the Good Thinking website. This project aims to provide access to online interventions targeting mental health of Londoners. For this project, I used the existing quality assessment I developed for this thesis (Study 1) and tailored it to online mental health interventions.

Third, I am using my knowledge in the Evaluation Exemplar project which aims to provide a guide on how to evaluate digital interventions. The target audience are developers of digital intervention who have no or little prior knowledge of embedding evaluation into their product development.

Fourth, this work has gained some interest from researchers across various disciplines. Specifically, I was invited to conduct a lecture sharing knowledge from my thesis to MSc students in Health Informatics for which I received excellent student feedback. My research has also led to collaborations with researchers at the University of Milano-Bicocca and National University of Singapore.

Specifically, as outlined in this thesis, currently there is no one set of standards aimed at supporting the development of digital behaviour change interventions. Development of a set of standards is a step forward to ensuring that users have access to digital tools that are based on evidence and guidelines. Lastly, my plans are to pursue in obtaining funding for the next stage of this research which will be a direct continuation of the studies reported in this thesis. These are outlined in section 'Immediate future research direction' in the Discussion Chapter and will include a definitive crossover trial of two popular PA apps as well as a strategic behaviour analysis of the current PA interventions implemented in the UK.

## Table of contents

<b>Declaration</b> .....	<b>2</b>
<b>Acknowledgements</b> .....	<b>3</b>
<b>Abstract</b> .....	<b>6</b>
<b>Impact statement</b> .....	<b>7</b>
<b>Table of contents</b> .....	<b>9</b>
<b>List of tables</b> .....	<b>11</b>
<b>List of figures</b> .....	<b>14</b>
<b>List of abbreviations</b> .....	<b>17</b>
<b>CHAPTER 1. Introduction</b> .....	<b>20</b>
1.1 Chapter overview .....	20
1.2 The importance of physical activity .....	20
1.3 Physical activity definitions, measurement and classification.....	21
1.4 The magic polypill .....	22
1.5 Meaningful change in physical activity.....	27
1.6 The current physical activity recommendations – every minute counts.....	27
1.7 Digital health .....	29
1.8 Digital social inequalities .....	30
1.9 Why mobile applications?.....	32
1.10 Focus on prevention and behaviour change.....	33
1.11 Physical activity policy environment .....	34
1.12 Prevention and technology.....	37
1.13 Theoretical framework and approach (the application of behaviour science tools).....	37
1.14 Aim and the structure of the empirical research of this thesis.....	42
<b>PART 1: FOCUS ON THE MARKET</b> .....	<b>44</b>
<b>CHAPTER 2. The quality of publicly available physical activity apps: a review and content analysis</b> .....	<b>44</b>
2.1 Chapter overview .....	44
2.2 Background.....	46
2.3 Methods .....	65
2.4 Results.....	74
2.5 Discussion.....	105
<b>CHAPTER 3. The relationship between popularity and app quality: regression models</b> .....	<b>114</b>
3.1 Chapter overview .....	114
3.2 Background.....	116
3.3 Methods .....	125
3.4 Results.....	137
3.5 Discussion.....	163
<b>PART 2: FOCUS ON THE USERS</b> .....	<b>175</b>
<b>CHAPTER 4. Methodological considerations: the feasibility and acceptability of a randomised crossover trial assessing two popular physical activity apps</b>	<b>176</b>
4.1 Chapter overview .....	176
4.2 Background.....	176
4.3 Methods .....	187

<b>CHAPTER 5. Results and discussion: the feasibility and acceptability of a randomised crossover trial assessing two popular physical activity apps ...</b>	<b>211</b>
5.1 Chapter overview .....	211
5.2 Results .....	211
5.3 Discussion.....	246
<b>CHAPTER 6. Process evaluation of a randomised crossover trial of two popular apps and identification of influences on physical activity: data-prompted interviews.....</b>	<b>262</b>
6.1 Chapter overview .....	263
6.2 Background - Methodological considerations .....	264
6.3 Methods .....	277
6.4 Results .....	281
6.5 Discussion.....	348
<b>CHAPTER 7. Discussion.....</b>	<b>363</b>
7.1 Chapter overview .....	363
7.2 Principal findings of the thesis .....	363
7.3 What is the public health potential of PA apps on the market? .....	364
7.4 Immediate future research direction .....	375
<b>References .....</b>	<b>383</b>
Appendix A: Apps used in the sample identification process (Study 1).....	407
Appendix B: Screenshot of the data extraction form (Study 1) .....	417
Appendix C: The manual for the extraction of data safety (Study 1).....	418
Appendix D: SUS descriptive record for both reviewers (Study 1).....	423
Appendix E: The BCT extraction Manual (Study 1) .....	438
Appendix F: Individual-level data for the sample of apps assessed (Study 1) .....	455
Appendix G: Directed Acyclic Graph (Study 2).....	472
Appendix H: Sample size calculation for the association between the number of BCTs and user ratings (Study 2).....	474
Appendix I: Table of different designs explored to assess PA app (Study 3) .....	475
Appendix J: Summary of the process of selection of correlates/determinants of PA to measure (Study 3) .....	479
Appendix K: Mapping of the determinants/correlates of physical activity (Study 3)....	481
Appendix L: Recruitment materials (Study 3) .....	483
Appendix M: Sensitivity analyses using vector magnitude (Study 3) .....	486
Appendix N: Results of the sensitivity analysis of self-reported physical activity variable log transformed (Study 3).....	487
Appendix O: Results of the sensitivity analysis for expected outcomes of physical activity (Study 3) .....	488
Appendix P: Topic guide for data-prompted interviews (Study 4) .....	489

Appendix Q: Behaviour Change Wheel matrices showing the links between different components of the Wheel .....	494
Appendix R: Published peer-reviewed articles .....	495

## List of tables

Table 1: Relevant definition relating to the behaviour addressed in this thesis .....	21
Table 2: The accepted MET classification for PA intensity: .....	22
Table 3: Relative risk reduction observed when comparing active and inactive participants .....	25
Table 4: The structure of the empirical research of this thesis .....	43
Table 5: Entertainment item of the MARS scale.....	59
Table 6: The summary of the healthcare quality indicators as applied to PA apps .....	65
Table 7: Data privacy and security assessment .....	68
Table 8: Five questions used in the review when assessing usability.....	71
Table 9: Descriptive data – iTunes.....	77
Table 10: Descriptive data – GP .....	78
Table 11: Descriptive data for the availability of the privacy policy (N=65) .....	85
Table 12: Descriptive data for the data privacy and security assessment (N= 46).....	85
Table 13: descriptive statistics for the inclusion of the BCTs.....	89
Table 14: Descriptive statistics for the app features .....	92
Table 15: Major feature group and subgroups mapped with BCTs.....	95
Table 16: Descriptive data for the quality of app development and evaluation process: organisational affiliation, expert and user involvement, and evidence of evaluation in peer-reviewed journals.....	99
Table 17: descriptive statistics for user ratings in iTunes .....	101
Table 18: descriptive statistics for user ratings in GP .....	102
Table 19: Descriptive data for the SUS assessment .....	103
Table 20: Descriptive statistics for the predictor: number of BCTs .....	138
Table 21: Descriptive statistics for the weighted star average of both stores (app level data) .....	139
Table 22: Descriptive statistics for the analysis of the confounders.....	142
Table 23: The analysis using univariate logistic regression models with the outcome of high versus low ratings. ....	143
Table 24: Odds ratios for the association between expert involvement and star average with expert involvement and app store interaction.....	150

Table 25: Odds ratios for the association between the number of app features and star average with number of app features and app store interaction .....	152
Table 26: Odds ratios for the association between usability and star average and app store interaction .....	153
Table 27: Odds ratios for the association between app store ranking for free apps and star average with app store interaction.....	154
Table 28: Odds ratios for the association between app store ranking for paid apps and star average with app store interaction.....	154
Table 29: Odds ratios for the association between the presence of peer-reviewed studies and star average with the presence of peer-reviewed studies and app store interaction .....	155
Table 30: The results of the t-tests comparing the number of BCTs between apps with and without evidence of expert involvement.....	156
Table 31: The results of the t-tests comparing the number of BCTs between apps with higher and lower usability.....	157
Table 32: The results of the t-tests comparing the number of BCTs and apps above and below the median size (iTunes).....	158
Table 33: The results of the t-tests comparing the number of BCTs and apps above and below the median size (GP) .....	158
Table 34: The results of the t-tests comparing the number of BCTs between apps with and without evidence of the presence of peer-reviewed studies.....	159
Table 35: Sensitivity analysis including BCTs with only high level of confidence in the presence in the apps (++) .....	161
Table 36: Sensitivity analysis assessing the association between BCT and star ratings (5 versus 1,2,3,4 stars) .....	162
Table 37: Differences between liked / popular and effective apps .....	166
Table 38: Primary and secondary objectives of Study 3 and 4 .....	187
Table 39: TIDieR template for the study interventions .....	192
Table 40: Summary of data collection at different data collection points.....	203
Table 41: Baseline characteristics of the sample (n=66) .....	213
Table 42: The 8 themes that emerged to the responses to the open-ended question: What do you expect from the apps?.....	218
Table 43: The type and the frequency of the PA apps downloaded prior to taking part in the trial.....	219



Table 44: Primary analysis – objective measure of change in PA (accelerometer): change in daily PA count (Axis 1 CPM), change in MVPA, time spent in light, moderate, vigorous PA, SB, step count, and wear time from baseline to Period 1 (in minutes)..	222
Table 45: The difference in change between App A and B using objectively measured PA including period effects .....	223
Table 46: the proportion of those participants that increased their MVPA for both periods.....	224
Table 47: Sensitivity analysis showing the primary analysis of the effect of the apps on those that engaged with the interventions for App A and B (in minutes).....	226
Table 48: Primary analysis - subjective measure of change in PA: Change in self- reported PA, ESE, exercise intentions, PA OE from post-baseline to Period 1 .....	229
Table 49: The difference in change between app A and B in self-reported PA, ESE, exercise intentions, PA OE including period effects .....	230
Table 50: Linear regression of the association between baseline and objectively measured PA and the change during the intervention period.....	231
Table 51: Descriptive statistics for the SUS score.....	235
Table 52: The frequency of SUS scoring for both apps with a score of 72.5 to mark the threshold for good usability .....	236
Table 53: Star ratings assigned to each of the apps .....	237
Table 54: Descriptive statistics for the experiential items of the DBCI Engagement scale (1-7) for App A.....	239
Table 55: Descriptive statistics for the experiential items of the DBCI Engagement scale (1-7) for App A.....	240
Table 56: Descriptive statistics for the amount of use component of the DBCI Engagement scale .....	241
Table 57: Descriptive statistics for the depth of use component of the DBCI Engagement scale .....	241
Table 58: The frequency of responses for each period .....	242
Table 59: Descriptive statistics of the mean mood for each of the study periods.....	243
Table 60: The constructs of the theoretical framework of acceptability.....	267
Table 61: Labels, definitions and examples of COM-B.....	273
Table 62: Labels, definitions and examples of TDF domains .....	274
Table 63: Profile characteristics of participants that took part in the interview (N = 20) .....	282
Table 64: Summary of the acceptability of the trial design and procedures with example quotes.....	283

Table 65: Summary table of the themes with TDF/COM-B classification for the domain Motivation .....	293
Table 66: Summary table of the themes with TDF/COM-B classification for the domain Opportunity .....	317
Table 67: Summary table of the themes with TDF/COM-B classification for the domain Capability .....	332
Table 68: A diagram of complimentary studies of this PhD and the next step proposed .....	377

## List of figures

Figure 1: Summary of the UK CMOs PA guidelines for adults and older adults.....	28
Figure 2: PAF for risk factors for all-cause years of life lost rate per 100 000 population for England, Scotland, Wales, and Northern Ireland, both sexes, 2016.....	34
Figure 3: COM-B model .....	40
Figure 4: the Behaviour Change Wheel with TDF domains COM-B model .....	41
Figure 5: Descriptive ratings representing the SUS score .....	71
Figure 6: Hypothetical examples of coders' agreement of 6 BCTs with a comparison of kappa and PABAK. ....	73
Figure 7: Flow chart of the apps included in the analysis .....	75
Figure 8: Examples of the features in the workout apps.....	80
Figure 9: Examples of the features in the tracking of running apps .....	81
Figure 10: Examples of the features in the Running programmes .....	82
Figure 11: Examples of the features in the pedometer-based apps group.....	83
Figure 12: Example interval timer app in Interval Timer by Deltaworks .....	84
Figure 13: the distribution for BCTs combined for free and paid apps. ....	89
Figure 14: Percentage of apps that included BCTs, presented by groups .....	91
Figure 15: Frequency of individual BCTs within the groups BCTs .....	92
Figure 16: Distribution of user ratings in iTunes and GP .....	102
Figure 17: Total SUS score.....	105
Figure 18: The visual classification of the significant associations found in studies that assessed user ratings .....	122
Figure 19: The visual classification of the significant associations found in studies that assessed the inclusion of BCTs .....	122
Figure 20: The histogram of the star rating distribution .....	128
Figure 21: Scatter plot with linear regression of the pilot data (N=10) .....	129

Figure 22: Data clustering in the current study .....	136
Figure 23: Distribution of star average for both stores.....	139
Figure 24: star average distribution between the pilot data (N= 10) and full sample (N=64) .....	140
Figure 25: The graph showing the results of the primary analysis .....	144
Figure 26: Star ratings for the same apps across the app stores (n= 17) .....	146
Figure 27: The primary analysis of the number of BCTs with user ratings with the addition of an interaction between the store and the number of BCTs .....	147
Figure 28: Frequency of individual BCTs within the two Groupings of self-regulatory BCTs .....	148
Figure 29: Odds ratios for the associations between BCTs that have been shown to be effective and a 4- or 5- star rating .....	149
Figure 30: Logistic regression showing the effect of the expert involvement on user ratings (expert involvement x store interaction).....	151
Figure 31: Logistic regression showing the effect of the number of app features on user ratings (number of app features x store interaction) .....	152
Figure 32: Logistic regression showing the effect of usability on user ratings (usability x store interaction).....	153
Figure 33: Logistic regression showing the effect of the presence of peer-reviewed studies on user ratings (the presence of peer-reviewed studies x store interaction)..	155
Figure 34: The results of linear regression between the store rank and the number of BCTs .....	160
Figure 35: the overall structure of this sequential mixed-methods study .....	175
Figure 36: The process of negotiating the design for this study.....	177
Figure 37: Randomised crossover trial design used in this study .....	179
Figure 38: Study design schema (phase 1).....	188
Figure 39: Flowchart for recruitment and retention to the crossover trial .....	212
Figure 40: Word cloud representing the most frequently words used by the participants to describe their expectations of the app interventions.....	217
Figure 41: distribution of change in MVPA showing the proportion of participants who increased, decreased and did not change their PA (Period 1).....	225
Figure 42: Scatterplot with a regression line displaying the relationship between baseline SB and change in SB.....	232
Figure 43: Scatterplot with a regression line displaying the relationship between baseline SB and change in PA.....	233

Figure 44: Graph showing the results of the assessment of potential mediating variables on the likelihood of achieving 20% increase in MVPA .....	235
Figure 45: The distribution of SUS for App A and App B .....	237
Figure 46: Distribution of user ratings for App A and B.....	238
Figure 47: Visualisation of the experiential items score for App A .....	239
Figure 48: Visualisation of the experiential items score for App B .....	240
Figure 49: The distribution of the mean mood for all 3 periods .....	243
Figure 50: The distribution of the change in mood variable .....	244
Figure 51: Scatterplot with smoothing spline showing the relationship between mood and PA count .....	245
Figure 52: Key functions of process evaluation and relations among them .....	265
Figure 53: sample screenshots of the usage of the apps used in the interview .....	269
Figure 54: sample of the multimedia diaries provided by the participants used in the interview: a) notes and screenshots, b) photos .....	270
Figure 55: Summary of the themes with COM-B / TDF classification .....	292
Figure 56: The <i>advise</i> stage relating to behaviour change in the 'ask-assess-advise' approach.....	371
Figure 57: Social ecological model of active living.....	374

## **List of abbreviations**

App – Application

AS – April Slee

BCT – Behaviour Change Technique

BCW – Behaviour Change Wheel

CI – Confidence Interval

CMOs' - Chief Medical Officers'

COM-B – 'Capability', 'Opportunity', 'Motivation' – 'Behaviour'

CONSORT – Consolidated Standards of Reporting Trials

CPM –Counts Per Minute

CWIS – Cycling and Walking Investment Strategy

DAG – Directed Acyclic Graphs

DBCI – Digital behaviour change interventions

DCMS – Digital, Culture, Media and Sport

ESE – Exercise self-efficacy

EM – Elizabeth Murray

EMA – Ecological Momentary Assessments

FH – Fiona Hamilton

GA – Ghadah Alkhalidi

GF – Gillian Forbes

GPPAQ – General Practice Physical Activity Questionnaire

GP – Google Play

HIIT – high intensity interval training

ICO – Information Commissioner’s Office

IPAQ – International Physical Activity Questionnaire

ITT – Intention-to-Treat

IQR – Interquartile Range

JITAI – Just In Time Adaptive Intervention

NHS – National Health Service

LA – Lou Atkins

MARS – Mobile Application Rating Scale

MVPA – Moderate to vigorous physical activity

MET – Measures oxygen consumption

MRC – Medical Research Council

NICE – National Institute for Health and Care Excellence

OTA – Online Trust Alliance

OE – Outcome expectancy

PA – Physical activity

PABAK – Prevalence and Bias Adjusted Kappa

PAF – The population attributable fraction

PB – Paulina Bondaronek

PHE – Public Health England

PII – Personally Identifiable Information

PIS – Participant information sheet

PPI – Patient and public involvement

RCT – Randomised Controlled Trial

RoB – Risk of Bias

SB – Sedentary behaviour

SD – Standard deviations

SUS – System Usability Scale

TDF – Theoretical domains framework

TIDieR – Template for Intervention Description and Replication

TIPPME – Typology of interventions in proximal physical micro-environments

UCL – University College London

UK – United Kingdom

YLLs – Years of life lost

## **CHAPTER 1. Introduction**

### **1.1 Chapter overview**

In this chapter I provide the background to the thesis by introducing the definitions and measurement of PA, by emphasizing the importance of activity to physical and mental health, and by outlining the impact of physical inactivity in the UK and worldwide. I then describe the current PA recommendations but also present evidence that even small increases in PA can have a meaningful impact. Next, I outline the rapid expansion of digital health technology, and provide a rationale for my focus on PA mobile apps specifically. I then describe the current PA policy environment as well as the current policy focus on prevention and technology, and provide the overview of the theoretical underpinnings to this thesis. I finish with highlighting the research gaps and present the overall aim and structure of this thesis.

### **1.2 The importance of physical activity**

Physical inactivity is an established independent risk factor for a range of serious health conditions including cardiovascular disease, diabetes mellitus and cancer [1-3]. In contrast, PA is associated with marked reduction in the risk of premature mortality and reduction in the risk of over 20 long-term conditions [4]. PA is also associated with improved mental health [5, 6]. Most international PA guidelines, including the UK, recommend 150 min of moderate or 75 min of vigorous intensity PA. Participating in muscle-strengthening activities (e.g. exercising with weights, carrying shopping bags) are also recommended twice weekly along with reducing the amount of sedentary time [7, 8].



### 1.3 Physical activity definitions, measurement and classification

*Physical activity* and *exercise* terms are sometimes used interchangeably despite their differences in meaning. Below are the definitions relation to PA (Table 1). In this thesis, I will focus on PA and, less frequently exercise, as the behaviours addressed.

Table 1: Relevant definition relating to the behaviour addressed in this thesis

PA	Any bodily movement produced by skeletal muscles that results in energy expenditure [9]
Exercise	A subset of PA that is planned, structured, and repetitive and has as a final or an intermediate objective the improvement or maintenance of physical fitness [9]
Sport	A subset of exercise that can be undertaken individually or as a part of a team whereby participants adhere to a common set of rules or expectations, and a defined goal exists [10]
Physical fitness	A set of attributes that are either health- or skill-related [9]

A standard measure of PA is a metabolic equivalent of task (MET) which measures oxygen consumption, i.e., the cost of conducting a specific PA. One MET is considered a resting metabolic rate, i.e., the amount of oxygen consumed at rest, and is equivalent to a caloric consumption of 1kcal/kg/hour. Table 2 shows the accepted MET cut off points for intensity of PA. A compendium of Physical Activities [11] includes 821 activities with MET scores used for quantifying the types of PA.

Table 2: The accepted MET classification for PA intensity:

Intensity	Quantification using METs [11]
Sedentary behaviour	1.0–1.5 METs
light-intensity	1.6–2.9 METs
moderate-intensity	3–5.9 METs
vigorous-intensity	≥6 METs

## 1.4 The magic polypill

“If PA was a drug, we’d talk about it as a miracle cure” UK Chief Medical Officers [7]

As described above (section 1.3), PA and exercise are often used interchangeably. It has to be noted that studies concerned with assessing the benefits of PA often do not distinguish between structured exercise as a subset of PA, and PA. Yet, the benefits of PA accrue through exercise and non-exercise activity. In this section, I describe the research of the impact of PA.

### Immediate effects of a single dose PA

There is evidence for the immediate positive impact of PA. For example, a single session of moderate-to-vigorous PA can decrease anxiety and insomnia, and improve cognitive function [12]. The observed short-term benefits of PA include lowering of blood pressure, reduction in glucose and lipids, and muscular relaxation. However, these changes only last a few hours [13]. Hence, regular PA is needed to produce sustainable health benefits. The processes by which such health benefits are achieved are termed chronic adaptations, and occur as the body responds to the demands of the activity by making specific changes.

## **PA adaptations**

PA adaptations have a positive impact on the sympathetic nervous system and the hypothalamic-pituitary axis (important in response to stress), and manifest by increased resilience to both physical and mental stress. The changes produced by regular PA that protect against long-term conditions, such as cardiovascular disease [14, 15], include reduction of visceral adiposity associated with impaired glucose and lipid metabolism. PA also reduces systemic inflammation [16] linked to metabolic syndrome and insulin resistance [17]. Regular PA also has an impact on other inflammatory diseases, such as rheumatological diseases [18, 19], respiratory diseases [20, 21], and cancer [22, 23]. In addition, it has an effect on the brain function as it increases synaptic plasticity and supports the underlying system involved in neuroplasticity including neurogenesis, metabolism and vascular function [24]. Thus, by having beneficial effects on learning and memory, it has been shown to protect against cognitive dysfunction and to alleviate depression [25], comparable to the effects found with antidepressant treatment [26].

## **Benefits of PA on physical and mental health**

The consequent experienced benefits of PA relate to both physical and mental health. The physical health benefits of PA are well established and include a lowering the risk of cardiovascular disease, hypertension, and diabetes mellitus [27, 28]. As obesity is linked to cancer [29], PA can help to maintain a healthy weight [30]. Moreover, physical inactivity is associated with carcinogenesis through a pathway independent of obesity [31]. Hence, PA is protective against cancer, most notably for breast and colon cancer [29] and evidence for the role of inactivity has been establishing for other types of cancer [32]. The effects of PA on mineral bone density are strong [33] and hence regular PA protects against osteoporosis [34]. Through improved balance and strength, PA reduces

the risk of falls and increases the time in independent living for older adults [35]. The summary of physical health benefits on PA are summarised in Table 3.

Table 3: Relative risk reduction observed when comparing active and inactive participants (reproduced with permission from [36])

<p><b>All-cause mortality</b></p> <ul style="list-style-type: none"> <li>• 31% risk reduction</li> <li>• 45% risk reduction when aerobic fitness is assessed</li> </ul>
<p><b>Cardiovascular disease</b></p> <ul style="list-style-type: none"> <li>• 33% risk reduction</li> <li>• 50% or greater risk reduction when aerobic fitness is assessed</li> </ul>
<p><b>Stroke</b></p> <ul style="list-style-type: none"> <li>• 31% risk reduction</li> <li>• 60% or greater risk reduction when aerobic fitness is assessed</li> </ul>
<p><b>Hypertension</b></p> <ul style="list-style-type: none"> <li>• 32% risk reduction</li> <li>• 50% or greater risk reduction when aerobic fitness is assessed</li> </ul>
<p><b>Colon cancer</b></p> <ul style="list-style-type: none"> <li>• 30% risk reduction</li> </ul>
<p><b>Breast cancer</b></p> <ul style="list-style-type: none"> <li>• 20% risk reduction</li> </ul>
<p><b>Type 2 diabetes</b></p> <ul style="list-style-type: none"> <li>• 40% risk reduction</li> <li>• 50% or greater risk reduction when aerobic fitness is assessed</li> </ul>
<p><b>Osteoporosis</b></p> <ul style="list-style-type: none"> <li>• Bone adaptations to exercise are load dependent and site specific</li> <li>• Routine physical activity is associated with improved bone health</li> </ul>

For mental health, benefits have been shown for depression, [37-39] and anxiety [40, 41]. In addition, PA helps with recovery from drug addiction [42, 43]. Lastly, PA has an impact on cognition in healthy population and in people with neuro-degenerative diseases [44]. Specifically, benefits have been identified for cognitive performance in various ages [45], for cognitive improvements in multiple sclerosis [46], and memory function in older adults [47].

It has to be considered that the current understanding of the impact of PA on health is based on observational studies. Such study design has its limitations. First, the issue of causality is a constraint. Specifically in PA research, if healthy people are more likely to exercise, the effect is reversing the cause and effect. Second, there may be confounding factors that are unmeasured. Third, although the use of objective measures over subjective measures is recommended, both have its limitations. There are issues with recall when using self-reported measures, especially when recalling longer periods. Although objectively measured PA eliminates the issues with recall, adherence to wearing accelerometers is an issue. Hence, accelerometry is often used for a limited time only to increase wear time. However, this introduces a problem with generalisability of the measured period. In addition, some accelerometer devices are better at recording different intensity of PA and hence might not be sensitive enough to detect activity.

### **Adherence to PA recommendations**

Adherence to PA recommendations is associated with 20–30% risk reductions for various long-term medical conditions and premature mortality [48]. However, In the UK 33% of men and 45% of women do not achieve the recommended level of PA and the prevalence of physical inactivity is well-documented [49].

Worldwide, insufficient PA is one of the 10 leading risk factors for global deaths, causing 3.2 million deaths each year [50], and the levels of inactivity are rising. For example, an analysis of data over 15 years found that between 2007 and 2016, total daily sitting time for adults and adolescents increased by roughly one hour [51]. One in four adults and more than 80% of the adolescent population around the world are not physically active enough [52]. In the UK, 29% of people are inactive (which is defined as not achieving at least moderate intensity PA per week) [53]. When considering non-participation in sports, the rate is higher: 57% [54]. Physical inactivity leads to 17% of premature deaths in the

UK [55]. At the same time, 70% of adults report they would like to increase their PA levels [56].

## **1.5 Meaningful change in physical activity**

A growing evidence base exists indicating that even lower doses of PA than the ones recommended (moving from inactive state) show the highest relative benefits. Arem and Irwin [57] showed that engagement in any PA is associated with lower risk of mortality with even small increases in exercise levels producing a significantly beneficial impact on health. Most recently, a systematic review of current systematic reviews provided strong evidence for curvilinear relationships between PA and health outcomes with health benefits observed with minor PA [58]. This means that change from an inactive to mild or moderately active shows relatively large risk reduction whilst further increase in PA volume yields relatively smaller risk reduction. For example, 0.1–3.74 metabolic equivalent (MET-h/week) of leisure time moderate to vigorous PA (MVPA) equating to 75 min/week of brisk walking) showed 1.8 year of life gained [59]. This finding has an important implication for public health suggesting that even small increases in PA can confer meaningful health outcomes, and that the biggest relative benefits come from engaging in light and moderate PA. As a result of this evidence, the latest UK CMOs PA guidelines [7] overturned the previous guidance that at least a 10 min bout of PA is required, and emphasized the message that any PA is beneficial.

## **1.6 The current physical activity recommendations – every minute counts**

The new UK Chief Medical Officers' (CMOs') PA guidance [7] published in September 2019 reflect the new evidence that more PA is better (beyond the recommendations, Figure 1).

Figure 1: Summary of the UK CMOs PA guidelines for adults and older adults (Source: [7], licensed under the Open Government Licence v.3.0)



Considering the prevalent low levels of PA and the new PA recommendation with a slogan “every minute counts” (Figure 1), it was considered that the advances in the field of digital health (in particular behaviour change) may help tackle high levels of inactivity [60, 61], and therefore may be considered as public health intervention if taken up at scale [62]. This chapter will now turn to discuss why digital health was considered as the



topic of investigation, and, specifically, why mobile apps were selected as the potential mode of delivery for public health interventions.

## **1.7 Digital health**

Digital technologies for health and healthcare enable the users to access and generate information about health and illness, engage in self-care practices and methods to promote health and well-being, and connect with other users and patients. These functions of digital health can be achieved via various means, such as conducting online searches for health and medical topics, blogs about health and medical issues, creating online support groups and communities, interactive online interventions, mobile apps, or wearables [63].

Health information seeking through websites was the primary function of early digital health. Still, accessing health-related information is a major motivation for online behaviour. For example, one of the top medical websites, WebMD attracts around 80 million visitors per month [64].

With the rapid development of digital technology, traditional websites and online discussion forums have been supplemented by social media platforms, which are now frequently used by patients with specific conditions to communicate with each other, and to share information and experience on specific health or medical topics of interest [65-67].

Telemedicine, defined as “medicine practiced at a distance”, originally using telephones, is increasingly incorporated into wireless devices, mobile apps, and data analytics [68]. In the UK and some parts of Europe, telemedicine has become a substantial contributor to healthcare delivery. In recent years, telemedicine has started to incorporate a range

of digital devices for patient self-monitoring at home, for example, attachments for “smartphones” which include sensors measuring blood pressure, heart rate, body temperature, etc.

Furthermore, digital devices can now be integrated with the body and the environment in various ways. Examples are versatile and include surgically implanted glucose monitors and insulin pumps, digital medication to increase adherence to a medication regime, mattresses and chairs that can detect movement in elderly in care to monitor their well-being, as well as robotic care providers.

Moreover, the opportunity for online health practices has been expedited by the introduction of “smartphones” and tablet computers with the pioneering iPhone release in 2007, Android in 2008, and the first iPad in 2010 [68]. In addition, wearables, i.e., self-tracking devices and sensors that are embedded in watches, wristbands, headbands, and woven into clothes to monitor (and often share) a multitude of variables, such as body weight (often enabling calculation of BMI), mood, body temperature, breathing and heart rate, geolocation, sleep patterns, body movement, etc.

## **1.8 Digital social inequalities**

The term digital social inequalities relates to the strong evidence that those with lower levels of education and income, older adults, those living in rural regions, as well as people living with disability, are less likely to participate in the Internet and digital device activities. The reasons for the lower likelihood of those social groups to go online include lack of basic digital literacy skills and experience, low interest or anxiety about learning a skill, and lack of access to the Internet or a device [69].

Availability of internet and online services has increased substantially with 90% of households having internet access and 89% of people going online on a weekly basis.

When considering mobile phone usage, “smartphones” have become the most popular internet-connected device with 78% of UK adults owning a device in 2016, in comparison to 39% in 2012 [70].

However, access to the Internet has improved for most of social groups, the UK digital exclusion heatmap [71] shows that, in the UK, low levels of Wi-Fi adoption correlate with lower levels of education, economic growth, and poorer health status. For example, 25% representing 3.5 million of people in the UK who are registered as disabled are not online [72].

In addition, there is evidence from the US that the “smartphone” owners are more likely to be younger, more affluent and have higher education, although no differences between ethnicity and racial groups currently exist [73]. There are no identified gender differences in relation to access and the use of the Internet and mobile devices in the countries of the Global North [74] although there are differences in how digital technologies are used. Women, for example, install 40% more apps of any type than men [75]. Men use slightly more health and fitness apps. Women are more likely to use diet and calorie counter apps [76]. In a more recent evidence from a sample of Dutch population, similar results were found with men using more fitness apps, and women using more use nutrition, self-care, and reproductive health apps [77]. These findings reflect the prevailing sociocultural gender differences where men are more likely to engage in PA as a socially acceptable activity to influence health, identity and body image, whilst women engage in diet-based activities related to their body image and expectations [78].

Hence, although some differences in usage pattern may be present, mobile phones have become ubiquitous and integral to people's lives in the UK. Because of the omnipresent nature of mobile devices, they provided novel avenues for behaviour change.

## **1.9 Why mobile applications?**

Within the new digital healthcare landscape, the rise of apps creates novel prospects for prevention opportunities and disease management [60]. Mobile health apps, as opposed to traditional face to face interventions, are more accessible [61], and provide a range of technology enhanced features, such as accelerometers, visualizations, tailored feedback and reminders. In addition, recent data show that mobile phone access is now as high among ethnic minority groups in higher income countries as in the rest of the population [79] and the use of mobile phones is increasing steadily in older populations [80], decreasing concerns about the effect of the digital divide on health inequalities. Hence, behaviour change interventions delivered using mobile health apps could have the potential to reach a large proportion of the population, thus increasing the public health impact of their small effects [62].

The mobile health app industry has doubled in the last three years, with around 325,000 health apps available in the major app stores at the end of 2017, and 3.7 billion health apps downloaded in the same year [81]. Sixty-five percent of apps in medical and health sections of the major app stores target exercise, diet and more generally wellness [68]. Hence, they potentially create a fertile landscape for prevention efforts. At the same time, in 2016, the NHS announced the plan to give patients free access to health apps for self-management of long-term conditions [82].

However, despite the wide distribution and popularity of health apps, many of them have been rapidly developed [83], and there is lack of evidence of their efficacy. For example,

a meta-analysis published by Direito, Carraça [84] found only 7 randomized controlled trials (RCTs) evaluating app intervention for PA and/or sedentary behaviour. Cumulatively, this study showed small effects of various mHealth interventions on PA and sedentary behaviours although the authors did not assess the impact of apps specifically.

### **1.10 Focus on prevention and behaviour change**

Non-communicable diseases resulting from behaviours as well as health inequalities resulting from the wider determinants of health place a significant burden on public finances. There is evidence that prevention initiatives show healthcare savings [85].

An analysis using Global Burden of Disease data [86] within the UK showed that all countries of the UK could reduce the burden of disease through effective prevention. Figure 2 illustrates this burden powerfully. The population attributable fraction (PAF) represents the proportion of years of life lost (YLLs) that could have been avoided if the exposure had been minimum in that year. Of the 10 highest risk factors, 8 were behavioural and metabolic factors. However, as the authors explain, the division between behavioural and metabolic risk factors is not absolute, because behavioural factors affect metabolic factors, such as high blood pressure and body-mass index. Indeed, low PA level is one of the top 10 risk factors. The multiple metabolic risk factors included in the list can be minimised by regular engagement in PA, as discussed in section 'The magic polypill'.

Figure 2: PAF for risk factors for all-cause years of life lost rate per 100 000 population for England, Scotland, Wales, and Northern Ireland, both sexes, 2016 (by [86], licensed under CC BY 4.0)

Rank	England	PAF (%)	Scotland	PAF (%)	Wales	PAF (%)	Northern Ireland	PAF (%)
1	Tobacco	19.26	Tobacco	22.76	Tobacco	20.31	Tobacco	20.01
2	Dietary risks	14.41	Dietary risks	16.12	Dietary risks	16.35	Dietary risks	15.88
3	High blood pressure	13.04	High blood pressure	14.62	High blood pressure	15.53	High blood pressure	14.99
4	High body-mass index	9.57	Alcohol and drug use	12.98	High body-mass index	9.85	Alcohol and drug use	11.50
5	Alcohol and drug use	9.52	High body-mass index	10.70	Alcohol and drug use	9.59	High body-mass index	9.97
6	High total cholesterol	7.44	High total cholesterol	8.49	High total cholesterol	8.07	High total cholesterol	8.35
7	Occupational risks	4.85	High fasting plasma glucose	5.02	High fasting plasma glucose	5.20	High fasting plasma glucose	5.18
8	High fasting plasma glucose	4.84	Occupational risks	4.63	Occupational risks	4.55	Occupational risks	4.30
9	Air pollution	4.04	Air pollution	3.87	Air pollution	3.91	Air pollution	3.58
10	Low physical activity	2.16	Impaired kidney function	2.48	Low physical activity	2.03	Low physical activity	2.52

■ Behavioural  
■ Environmental and occupational  
■ Metabolic

## 1.11 Physical activity policy environment

This section provides an overview of the PA policy in the UK focusing on the critical analysis by Milton & Bauman published in 2015, and outlines the developments since then.

In their critical analysis of the PA policy in England, Milton and Bauman [87] pointed that there has been an increased focus on the national policy to address PA levels with some important achievements.

More specifically, the approach to national PA recommendations in England demonstrated scientific robustness. In 2004, the landmark recommendations *At Least Five a Week* [88], included the CMO's PA guidelines. Based on 40 national and international experts review, *Start Active, Stay Active: A Report on Physical Activity from the Four Home Countries Chief Medical Officers* [89] updated the PA recommendations and included guidelines for children and older adults, and conveyed the risks of sedentary behaviours. However, the document was criticised for the lack of an implementation plan, and only two years later, 80% of the public could not remember the

current PA guidelines [89]. It should be considered that the intended target audience included policy-makers and health professionals.

*Everybody Active, Every Day* was Public Health England's (PHE) attempt to propose an implementation framework for the *Start Active, Stay Active* guidelines. This document has been influential in shaping both national and international policy [90].

The new UK PA guidelines were published in September 2019 to reflect evidence updates [7]. This coincides with the release of similar guideline updates across the globe (e.g. Canada, South Africa, USA) as well as the global action plan on PA [91]. There is an increasingly clear, coherent structure from international and national policy, but as yet the challenge of creating aligned local policy and implementation programmes remains.

Further, Milton and Bauman [87] identified important steps for development, specifically the need for a consistent and sustained surveillance system, and realistic short- and long-term national PA targets. Moreover, a critical analysis of the two large-scale campaigns in England, Active For Life and Change4Life, showed that these population-level interventions were not sufficient to elicit sustainable behaviour change on a population level [87]. The authors argued that the length of these campaigns was too short to produce an effect.

Since Milton and Bauman [87] critical analysis, the main surveillance measure has changed from the *Active People* survey to the *Active Lives* survey. However, knowing the shortcomings of the *Start Active, Stay Active* guidelines, the new guidelines are accompanied by communication and surveillance expert working groups to address some of these challenges. This work remains ongoing.

There is some argument that when the first surveys are completed and the prevalence against the new guidelines are assessed, population PA will likely seem to rise, given the more flexible nature of the guidelines. However, this is unlikely to be the case in reality; the surveillance goalposts have simply changed and this may make the UK's policy efforts seem more effective than they actually are.

*Sporting Future - A New Strategy for an Active Nation* [92] published in 2015 by the Department for Digital, Culture, Media and Sport (DCMS) included key actions and mandates for government and the wider PA sector. This strategy was seen as a reversal of the government emphasis during the Olympic build-up and legacy periods which focused on sport participation. This document emphasises the key potential societal benefits of PA (referred to as the five pillars).

In 2016 Sports England was charged with implementing the strategy towards an active nation. Alongside the aforementioned Active Lives survey that was developed for the first time, Sports England placed emphasis on general PA and their funding priorities reflected the five pillars, rather than participation levels per se.

In 2017, *Cycling and Walking Investment Strategy* (CWIS) for England was published [93]. This is the only framework with a statutory bind, i.e., it includes targets that are required by law to be reached.

In summary, the UK's primary policy response to inactivity is the production of evidence-based guidelines. A coherent structure from international and national policy emerges, but consistent implementation strategies still remain a challenge. In addition, the mass public health campaigns targeting PA show limited effectiveness [94-96].



## **1.12 Prevention and technology**

The emphasis on prevention and realising the potential of technology for health cannot be underestimated in the current health politics. Two of the three main priorities of the current Secretary of State for Health and Social Care are advancing health technology and prevention [97]. Hence, this thesis is timely as it is contributing to the knowledge of using digital technology to improve health and well-being, and prevent disease. However, prevention requires behaviour change. Duncan Selbie, Chief Executive of PHE has emphasized the importance of behaviour science:

*“The behavioural and social sciences are the future of public health. The evidence we now have from behavioural science can help us understand and therefore influence behaviour change that promotes health, prevents disease, and reduces health inequalities. We must reach and be meaningful to people in the lives that they are leading.”*

*In Improving people's health: Applying behavioural and social sciences to improve population health and wellbeing in England [98, p.4 ]*

In line, the theoretical underpinnings of this thesis are based in health psychology, specifically the recent advancements in the field of behaviour change. The following section of this Introduction focuses on the behaviour change framework that was considered as the theoretical foundation for this project.

## **1.13 Theoretical framework and approach (the application of behaviour science tools)**

Models of social and health behaviours have been developed to explain the process underlying behaviour, and to predict and subsequently change behaviour, e.g., The

Health Belief Model [99], Social Cognitive Theory [100], Transtheoretical Model [101].

Two pertinent issues relating to the field of health psychology/behaviour change are considered in the context of the selection of the theoretical underpinnings of this thesis: 1) the limitations of the models to change behaviour, 2) the lack of guidance on the selection process of the most appropriate model.

Firstly, many of the frequently used models of behaviour rely on intention as the direct influencing factor on the behaviour. For example, the Theory of Planned Behaviour [102] states that the proximal determinant of volitional behaviour is the intention to engage in that behaviour. A meta-analysis of 47 experimental studies found that medium-to-large change in intention ( $d = 0.66$ ) produced only a small-to-medium change in behaviour ( $d = 0.36$ ) [103]. The authors also found that intentions had less impact on behaviour when there was less perceived behavioural control, there was higher potential for social reaction to the behaviour (relating to social acceptability of behaviour), and when the performance of the behaviour was conducive to habit formation (hence bypassing the intentions because of the higher automaticity of the behaviour).

In addition, the reliance on deliberate, rational decision-making has also come under scrutiny as models of social and health behaviours often do not address the influence of automatic motivation on behaviour, such as habits, emotions and impulses [104, 105]. Moreover, these models underestimate the influence of the environment on the behaviour. The quote from Theresa Marteau summarises the power of the *choice architecture* on the behaviour: “environments exert a stronger impact on what people do than what’s in their minds” [106, p.4].

Secondly, UK Medical Research Council (MRC) *guidance on developing and evaluating complex interventions* [107] advocate the use of theory to design complex interventions.

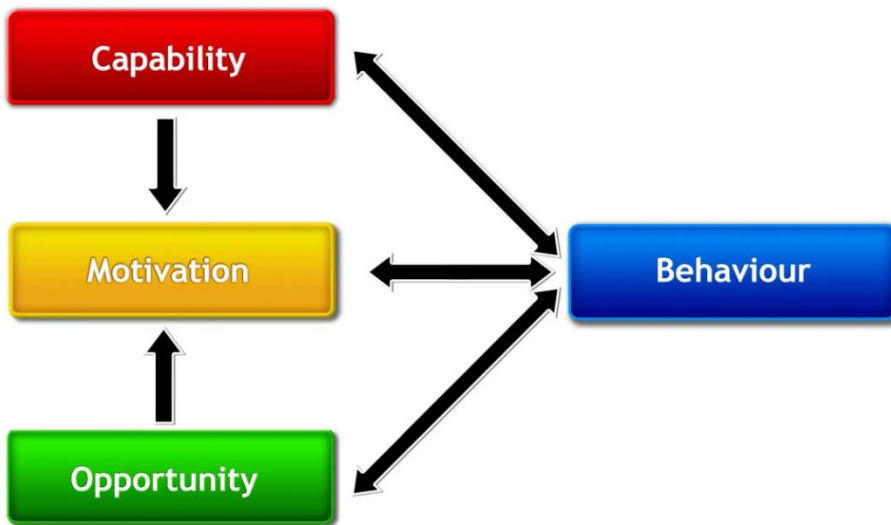
However, a systematic process of theory selection is lacking from the guidance or wider literature [108]. The high number of models of behaviour with many of their theoretical constructs overlapping means that comparing, synthesising, and ultimately progressing, the research on behaviour change intervention is challenging [109, 110].

### **The Behaviour Change Wheel**

Behaviour is a result of various interacting factors occurring in an ecological system [111]. Behaviour change interventions should be developed based on the understanding of the factors that affect the target behaviour to subsequently identify the intervention content to target these influences [112]. Behavioural theory and evidence based tools have been developed to facilitate such understanding.

The Behaviour Change Wheel (BCW) [113, 114] includes a series of tools to systematically develop behaviour change intervention. In addition, the relevant components of the BCW can also be used to describe and classify the content of the existing interventions. The BCW was developed based on a synthesis of 19 frameworks of behaviour change. At the centre of the Wheel is the COM-B model of behaviour which provides a conceptualisation of the components that need to occur in order for the Behaviour to be performed. These are Capability, Opportunity, and Motivation (Figure 3). Capability can be divided into *physical* (physical skills) and *psychological* capability (mental skills, knowledge); opportunity relates to the external factors: *physical* (the environment) and *social* opportunity (e.g., social norms); motivation includes *reflective* (active thought processes) and *automatic* (impulses, habits, emotional reactions) motivation.

Figure 3: COM-B model (by [114] licensed under CC BY 2.0)



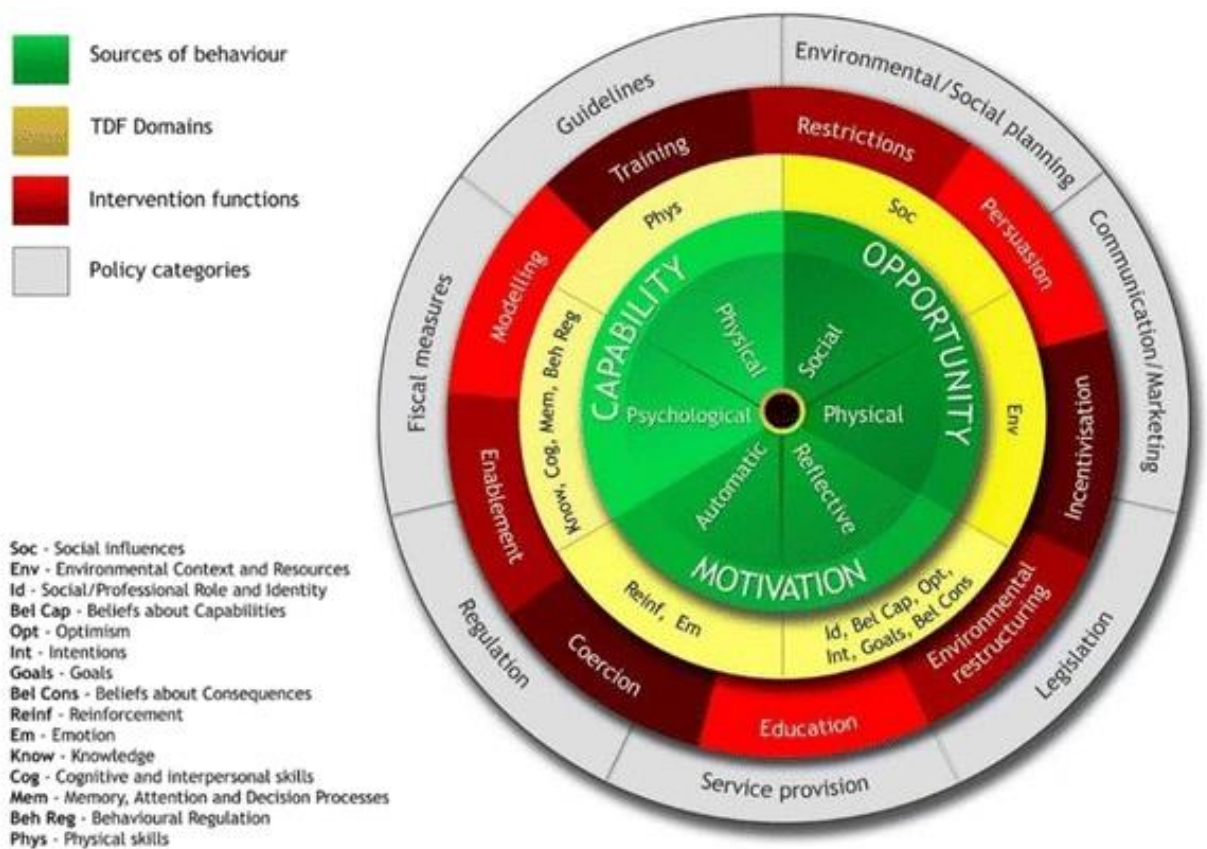
Capability and opportunity can influence both behaviour and motivation directly, for example, motivation to play a musical instrument may be increased by possessing the skill to play an instrument (capability) or by having access to an instrument (opportunity). Motivation can influence behaviour indirectly through capability and opportunity. For the behaviour to occur, motivation to perform the desirable behaviour needs to be stronger than the motivation to conduct other competing behaviours.

The COM-B model in the BCW is surrounded by nine intervention functions, i.e., the purpose the intervention has, these are: education, training, enablement, incentivisation, coercion, modelling, environmental restructuring and restriction. In turn, these functions are surrounded by seven policy categories, i.e., the channels through which the interventions are implemented. These are: guidelines, service provision, legislation, regulation, fiscal measures, communication and marketing, environmental and social planning.

## Theoretical Domains Framework

The second model that helps to understand and influence behaviour is the Theoretical Domains Framework (TDF) [115]. This framework is a synthesis of 33 psychological theories 84 psychological constructs. It consists of 14 theoretical domains: knowledge; skills; memory, attention and decision processes; behavioural regulation; social/professional role and identity; beliefs about capabilities; optimism; beliefs about consequences; intentions; goals; reinforcement; emotion; environmental context and resources; and social influences. The TDF maps onto the COM-B components. Figure 4 displays the BCW constituent parts with intervention functions, policy categories and shows how the TDF domains are linked to each COM-B component

Figure 4: the Behaviour Change Wheel with TDF domains COM-B model (by [114] licensed under CC BY 2.0)



## **Behaviour Change Techniques**

The absence of standardized definitions and labels for intervention components means that systematic reviewers often develop their own systems for classifying behavioural interventions to synthesize study findings [7–10]. This makes it difficult to replicate, synthesise the knowledge on the intervention effect in meta-analyses, and implement the interventions.

A 93-item Behaviour Change Techniques Taxonomy [116], developed through extensive international expert consensus, is a system of naming, describing and classifying the content of the behaviour change interventions. Behaviour Change Techniques (BCTs) are defined as the *active ingredients* in behaviour change interventions. They are identifiable, observable, replicable, irreducible components of an intervention designed to change behaviour.

In this thesis, I used the BCW and the BCT Taxonomy (v1) as frameworks that underpin this research work. Specifically, the BCT Taxonomy (v1) was used to i) describe and quantify the theory-based content of the popular PA apps on the market (Study 1), and ii) assess the association between the number of BCTs and user satisfaction (Study 2). The COM-B and TDF were used to iii) systematically map the correlates/determinants of PA to facilitate the selection of the measures for the feasibility trial (Study 3), and iv) categorise and describe in theoretical terms the influences, i.e., enablers and barriers to PA behaviours, identified in the interview following the trial (Study 4).

### **1.14 Aim and the structure of the empirical research of this thesis**

The overall aim of this thesis was to explore the public health potential of currently available health apps for increasing PA. I present the research work on this thesis in two distinguished parts. In Part 1 my focus is directed onto the app market to explore the questions about the quality of the popular PA apps that are available on the market. In

Part 2 my focus shifts onto the users to investigate the potential of the selected PA apps to change behaviour. This structure is illustrated in Table 4 below

Table 4: The structure of the empirical research of this thesis

<b>Parts</b>	<b>Studies</b>	<b>Aims of the constituent studies</b>	<b>Methods</b>
1: Focus on the market	Study 1	To ascertain the quality of current PA apps available on the market	Review and content analysis
	Study 2	To explore the relationship between the app quality indicators and popularity in current PA apps	Analysis using regression models
2: Focus on the users	Study 3	To investigate the feasibility of an RCT assessing two selected PA apps and to assess the effects of the app interventions on PA.	Feasibility crossover trial
	Study 4	To explore the acceptability of the trial, and to understand the issues around PA engagement and how apps can influence the PA engagement	Data-prompted interviews

## **PART 1: FOCUS ON THE MARKET**

In Part 1 of this PhD, I focused on the app market in order to address the questions of the quality of the most popular apps on the market. Specifically, I report on my investigation into the safety, likely efficacy, and the user experience of the most popular apps on the market.

### **CHAPTER 2. The quality of publicly available physical activity apps: a review and content analysis**

#### **2.1 Chapter overview**

**Introduction:** PA interventions delivered using mobile apps created novel prospects for public health interventions. Most of the current apps have not been evaluated and the quality of these apps is unknown. As there is no standard app quality assessment criteria, the approach of this study was to apply the healthcare quality indicators of safety, effectiveness, and provision of the best possible experience to develop an app quality assessment framework.

**Aim and objectives:** The aim of this review and content analysis was to evaluate the quality of the most popular PA apps on the commercial market. The objectives were to assess the safety, likely effectiveness and user experience of the sample of popular PA apps.

**Methods:** The top-ranked 400 free and paid apps in the Health & Fitness categories of iTunes and Google Play stores were screened for inclusion in the analysis. Apps were



included if the primary behaviour targeted was PA, targeted users were adults and the apps had stand-alone functionality. The apps were downloaded on mobile phones and assessed by two reviewers against the following quality assessment criteria: 1) users' data privacy and security, 2) presence of Behaviour Change Techniques and the quality of the development and evaluation processes, and 3) user ratings and usability.

**Results:** Out of 400 apps, 156 met the inclusion criteria and 65 apps were randomly selected to be downloaded and assessed. Nearly one-third (29.2%; 19 out of 65) of apps did not have a privacy policy. Most of the apps collected Personally Identifiable Information (PII) and shared users' data with a 3<sup>rd</sup> party. Every app contained at least 1 BCT, with an average number of 7, and a maximum of 13. All but 1 app had a commercial affiliation, 12 had consulted an expert and none reported involving users in the app development. Only 12 of 65 apps had a peer-reviewed study connected to the app. User ratings were high with only a quarter of the ratings falling below 4 stars. The median usability score was 86.3 / 100, which equates to "excellent".

**Conclusion:** Despite the popularity of PA apps available on the commercial market, there were substantial shortcomings in the areas of data safety and likelihood of effectiveness of the apps assessed. The limited quality of the apps may represent a missed opportunity for PA promotion.

## 2.2 Background

### 2.2.1 *The development of the app quality assessment tool*

The background to this chapter explains the process of developing the app quality assessment tool that was used to assess the PA apps on the market. The specific structure is outlined below:

- 1) the reviews that synthesised the concepts and measures used to assess quality of apps are described
- 2) the selection of app quality components relevant to PA apps is discussed
- 3) the operationalisation of each of the quality components is considered

#### **Concepts and measures used to assess quality of apps**

Quality is more than effectiveness, although there has been considerable debate about how exactly *app quality* should be defined, with a variety of frameworks available. Recent reviews by BinDhim, Hawkey [83] and Bardus, Smith [117] categorized and evaluated the methods used for quality assessment of apps.

The first review by BinDhim, Hawkey [83] demonstrated the variability in methods and measures used to review the quality of health apps. Some studies evaluated the quality based on downloaded app content, whilst others used the description of the apps available on the major app distribution platforms.

The methods used to assess the apps varied considerably and included use of scales adopted from websites' quality evaluation measures, presence of evidence-based content, quality assessment based on user ratings and reviews, and assessment of usability of app functions.

This review also showed inadequacies in the description of the quality assessment methods. The authors devised an eight-question checklist which described what information should be provided by reviewers of apps, such as the method of data collection and analysis, description of inclusion/exclusion criteria, country in which the app store was accessed, and a provision of a list of the identified apps. This tool helps to ensure that review methods for app quality evaluation are adequately reported and, in turn, replicated.

Similarly, in their review Bardus, Smith [117], found a variety of methods used to evaluate the quality of apps: out of 20 studies concerned with app quality assessment, 17 evaluated the content quality parameters, such as presence of evidence-based content and the use of theory. Four studies conducted formal evaluations of design quality parameters, such as usability and functionality. The approaches used to conceptualise and measure quality varied substantially, and the studies tended to focus on either the design quality or on the presence of evidence-based content, but not both. The authors called for more research to assess the quality of both design and content of health apps.

Both reviews were descriptive in their approach and did not provide any recommendations on the most important domains of app quality and what measures should be used to evaluate the apps in order to standardise the app review process. In absence of a standardised framework, this chapter will first consider the definition of quality in mobile health apps.

### **The selection of app quality components relevant to PA apps**

Health apps have the potential to be an important healthcare tool, hence healthcare quality indicators were considered appropriate to apply when assessing the quality of the

apps. The concept of quality in healthcare is complex and multifaceted [118]. Maxwell [119] proposed six dimension of healthcare quality: accessibility (ease of access to all patient groups), relevance to the need of the community, effectiveness, equity (fairness in the distribution), acceptability, efficiency and economy (desired health outcomes at the lowest cost). On the other hand, Donabedian [120] proposed a different categorization and argued for three crucial elements that pertain to the quality of healthcare: structure (facilities and healthcare professionals available), process (actions by which healthcare is provided), outcomes (the results of the actions).

The dimensions of quality proposed by Maxwell and Donabedian were developed before the existence of “smartphones” apps, and are perhaps more applicable to healthcare services provided face-to-face. As potential new healthcare tools, apps need a more concise approach, one that *High quality care for all: NHS Next Stage Review Final Report* [121] may provide. This report outlined the 10-year vision for the (National Health Service) NHS with strategies to improve the quality of care. In the Report, high quality healthcare was defined as being (1) safe, (2) effective and (3) providing the most positive experience possible. These quality indicators are simple yet comprehensive as they include broad aspects of healthcare. They are also sufficiently flexible to apply to potential new healthcare tools, such as PA apps. In the following section, I consider how the criteria of *safety*, *effectiveness*, and *provision of the best possible experience* can be applied to PA health apps.

## **The operationalisation of each of the quality components of PA apps**

### ***2.2.2 What does safety mean in health apps?***

When applying the healthcare quality indicator of safety, I argue that two aspects of safety needs to be considered: a) attempts to minimise the risk of injury or harm as a

result of using the PA apps, b) the security of users' personal information and their right to privacy.

Overall, more PA conveys greater benefits although higher intensity of PA may also increase risk of injury [122]. In order to reduce the risk of injury the American College of Sports Medicine recommends incremental increase in PA that is tailored to the individual fitness level, and stresses the importance of warm up, conditioning and/or strengthening, and cool-down. The factors that influence the risk of sustaining an injury from PA are diverse and include type and dose of activity, age, health status, equipment used, as well as environmental conditions.

In a review of research on adverse events associated with PA, it was found that the benefits of PA outweigh the risk of injury associated with PA [123]. In the light of these findings, I argue that 1) the benefits of potentially increasing PA through apps outweigh the risk of harm from PA apps, and 2) the factors increasing the risk of injury are complex and variable, and it would be challenging to measure the risk of injury from PA apps. Hence this quality indicator, albeit important, will not be assessed in this review.

The security of users' personal and health information, and their right to privacy, need to be considered when assessing the safety of health apps. Simple measures can be used to determine whether an app has basic levels of security and privacy, such as password protection, encryption, and explicit privacy policy [124]. Studies concerned with assessing security and privacy of online content would have used these indicators [125].

Yet, ensuring security of personal and health information proves challenging as manifested by the failure to identify insecure apps by a dedicated medical app store. The NHS App Library (closed in October 2015; recently reopened) was a database of health apps officially endorsed by the NHS. The process of peer review used in the Apps Library

ensured that apps were relevant for patients, used verifiable source of information, complied with Data Protection Act, and were clinically safe.

However, in a study of 79 apps accredited by The NHS Apps Library, which were all considered safe, it was found that 89% of the apps were breaching users' security by sending unencrypted information to online services, 66% of apps did not use encryption when sending users' personal and health information, and 20% did not have a privacy policy [60].

The authors of that study argued that regulatory bodies needed to establish more stringent rules for safety and security accreditation for the apps as, up until now, ensuring compliance with the principles of data protection relies extensively on developers [60]. Similarly, in their study on the security of apps, Wicks and Chiauzzi [126] argue that regulators are provided with insufficient resources to control the safety of an increasing number of apps, and that a new approach is needed, such as education of the users and pre-release expert reviews.

The study by Huckvale, Prieto [60] involved a 6-month analysis of the data transmissions between digital devices as well as an in-depth examination of the privacy policies of each app. This is beyond the scope of this review. However, apps with high quality should provide easily available and accessible privacy policy information [127]. The ePrivacy directive (2002/58/EC, as revised by 2009/136/EC [128]) states that access to user information is permitted after gaining the user's consent which can be given only after provision of "clear and comprehensive information" [Article 5(3)].

The Information Commissioner's Office (ICO) is the UK's independent regulatory body set up to uphold information rights. It covers legislation related to data protection and

privacy. In their guidance on privacy in mobile apps directed at app developers [127], the ICO recommends that information about users' data processing practices should be available before the app is downloaded. In addition, developer's transparency about the purpose of users' data gathering and sharing is crucial. A multi-lingual privacy statement is also recommended. Further, Online Trust Alliance (OTA) is a global organisation advocating best practices for security and privacy. In their guidelines, OTA advocates that a short form summary, in a simple jargon-free language, of the key points of the privacy policy should be made available to the potential user (with the full privacy policy available if the user wishes to access it). For the purpose of this study, the recommendations of the ICO and OTA were followed to design a tool assessing adherence to privacy and security of users' data.

### **2.2.3 What does “effective” mean in health apps?**

Scientific literature evaluating PA apps is in its early stage but some evidence of effectiveness emerges. Recently, Flores Mateo, Granado-Font [129] assessed the effects of interventions delivered using apps that aimed to reduce weight and increase PA. Meta-analysis of nine RCTs showed a significant decrease in body weight (-1.04 kg) and BMI (-0.43 kg/m<sup>2</sup>) compared to control group. The authors indicated a “nonsignificant difference in PA”. However, most of the studies focused on weight loss outcomes with only 5 of 12 studies including PA measure as secondary outcomes. As such, most of the app interventions focused on diet and nutrition. Hence, at the time of conducting these studies, there was limited evidence of PA app effectiveness that might relate to the lack of experimental evaluation research.

Although app development companies are obliged to provide evidence if they claim that their apps are effective, high quality data (e.g., from randomised controlled trials) is not required [130]. Thus, data on effectiveness of commercial apps is limited.

In the absence of conclusive results from research assessing effectiveness of apps, alternative methods of evaluation can be used. One approach is to assess the likelihood of effectiveness by a) evaluating the content of the app for adherence to evidence-based guidelines, and b) inclusion of behaviour change techniques.

In the UK, National Institute for Health and Care Excellence (NICE), an independent public body with a role to improve availability and quality of NHS treatments and care, produced evidence-based guidance on health behaviour change (NICE, 2014). McMillan, Hickey [132] analysed and converted NICE behaviour change guidance and applied it to apps in order to assess their adherence to evidence-based recommendations. The NICE recommendations were converted into a rating tool which was tested on apps from the NHS Apps Library. This study was the first that attempted to use NICE guidance to evaluate the health app market. One of the limitations of the tool is that, with 62 questions, the assessment is time-consuming. Secondly, some of the questions were difficult to evaluate without the input from the app developers. Thirdly, inter-rater agreement was poor for some of the questions, and each assessor interpreted some of the items in the tool slightly differently. Last, piloting was based on app description and not on the downloaded content. This is not recommended as the description of the app in the store might not be fully representative of the app content [83].

The scope of the NICE guidance is far-reaching, and it encompasses intervention developers, intervention deliverers, commissioners, as well as policy-makers. It also deals with formal assessment of effectiveness, and most of the apps on the market are not evaluated [133]. In this study, I argue, a more specific assessment tool addressing the quality of apps is needed to assess the likelihood of effectiveness of behaviour change apps aiming at increasing PA.



The second approach for assessing how likely the apps are to be effective is the inclusion of behaviour change techniques in the app content. The use of theory is vital in developing, evaluating and implementing effective behaviour change interventions [134]. However, choosing the most appropriate theory is challenging. There is an extensive body of work for classifying behaviour change interventions but none of the frameworks are comprehensive enough to cover a full array of intervention types, and there are often inconsistencies in the definitions of intervention components within these frameworks [135].

The development of the Behaviour Change Wheel [113] incorporates and classifies a wide range of behaviour change interventions and enables interventions to be developed using a systematic method by identifying what changes are needed, selecting appropriate intervention functions, and specifying intervention content using the Behaviour Change Technique (BCT) Taxonomy [116]. The BCT Taxonomy [116] has been used in previous studies that aimed to characterise BCTs in wide range of health apps tackling various types of behaviours, such as weight management [136, 137], alcohol reduction [138], and PA [139].

Research has been accumulating on evidence for effective BCTs for PA behaviour change. Studies show: PA self-monitoring is one of the strongest predictors of behaviour change [140]; goal setting is frequently used and effective in increasing PA [141]; individually tailored feedback is more effective than non-personalised generic information on PA [142]; participants' use of higher number of BCTs may be associated with better outcomes. Specifically, interventions where participants used more BCTs (11-15) was associated with higher increase in PA in comparison to those that used fewer BCTs [143]. In addition, NICE guidance for behaviour change (NICE, 2014) recommends utilising the BCTs taxonomy when developing behaviour change interventions.

The Taxonomy has predominantly been used to evaluate more traditional face-to-face interventions by extracting the BCTs from written materials, such as journal papers and intervention manuals [144]. Thus, assessing BCTs in the context of mobile technology can be challenging as researchers extracting the BCTs from the apps need to “translate” the observed features into BCTs [145].

In summary, there is limited evidence of effectiveness of PA apps. In the absence of high quality data from randomised controlled trials, there have been attempts to evaluate the likely effectiveness of PA apps by assessing their adherence to evidence-based guidelines and/or the use of behaviour change theories. For this study I utilised the BCT taxonomy to assess the potential for behaviour change of PA apps. In order to facilitate the process of assessing the apps, I mapped the app features and BCTs to help understand how BCTs are ‘translated’ in the apps. This will aid the extraction of the BCTs from apps in the future studies.

#### ***2.2.4 What does “providing the most positive experience possible” mean in health apps?***

Two important aspects of positive experience were considered in the context of app review: a) user ratings in the app stores, and b) measures of design quality.

##### **2.2.4.1 User ratings**

User ratings in app stores are an easily available measure that may provide valuable information on users’ perceptions of an app [146]. In a recent attempt to assess the association between user ratings and different characteristics of apps, Mendiola, Kalnicki [147] found that providing features that enable plan of action, tracking, the ability to send data to healthcare providers, and higher usability impacted positively on user ratings. However, the model explained only 9.3% of the ratings. This study used a general

sample of health apps but it is possible that user perceptions about app characteristics vary depending on the behaviour addressed by the app. More research is needed to assess what app characteristics and technology-enhanced features are associated with user ratings in order to shed light on user preferences for PA apps.

One problem with user ratings is that fake app ratings may be used to boost sales [148, 149] , and companies specialising in providing app reviews and ratings operate on the market, for example: <http://app-reviews.org>. Positive ratings can be bought but there is a reliance on genuine users of the app to mark it down if the app does not live to their expectations [150]. Despite this shortcoming, user ratings have two advantages: a) they are easily available and b) most users pay some attention to them, and, as such, including them adds to the external validity of the study, as it mimics user experience.

#### **2.2.4.2 Measure of design quality**

When defining ‘provision of the most positive experience possible’ in an app, the design aspects of an app needs to be considered. Users often choose an app based on perceived design quality and ease of use [130].

It is not surprising that design performs a salient role in eHealth interventions [151]. It plays an important role in addressing health inequalities in uptake and adherence to digital intervention through use of simple language and presentation of information in an audio-visual format [152]. However, there is no one standardised framework to assess the design quality of apps.

There is an inherent problem with assessing measures of app design quality as many of the design aspects are subjective in nature. Powell, Torous [153] assessed inter-rater reliability of 22 measures used to evaluate the quality of apps and found that most of the assessment tools had poor inter-rater agreement. In general, those that were more objective (assessed on facts) had high levels of inter-rater reliability, in comparison to

those that may be considered more subjective, such as ease of use. Nevertheless, design aspects, such as aesthetics, ease of use and tailoring play an important part in app uptake and usage, and any assessment of app quality need to address not only the quality of the content but also the quality of design.

Usability evaluation is a common method used in user-centred design to assess the quality of app design. There are various methods to assess usability, such as think aloud, where users comment on the product while using it; cognitive walkthrough, where task scenarios are used to assess any usability issues; and heuristic evaluation, which is typically conducted by the product designers. It involves evaluating the product according to the five usability factors: learnability, efficiency, memorability, error recovery, and satisfaction [154].

Jacob Nielsen's heuristic evaluation is one of the most popular tools for assessing usability is [155]. It enables identification of potential issues in the system that are likely to hinder its usability. This method is considered to be the standard for evaluation of user interface [156]. The advantages of the Nielsen's usability heuristics is that it needs fewer resources, in comparison with other methods of usability evaluation (such as think aloud or cognitive walkthrough, where the user is asked to perform certain tasks), and can be done by evaluators rather than users. It is often called a "discount" method.

However, the potential issue with the Nielsen's usability heuristics is that it consists of general usability principles which may prove difficult to reliably interpret. As the Nielsen's heuristics evaluation was developed almost 20 years ago for assessment of user interfaces in general, evaluators often need to tailor the heuristics to test new technologies. It is argued that for the purpose of the current study, a more specific tool for assessing the design quality of health apps was needed.

Most recently, Stoyanov, Hides [157] attempted to address the gap in the lack of specific measure in app quality assessment and developed the Mobile Application Rating Scale (MARS) which incorporates quality dimensions that are of importance to both health professionals and developers. The authors conducted a review of studies assessing quality across diverse research areas, such as mHealth and User Experience and extracted quality indicators. Subsequently, an expert panel comprising of health professionals, and digital technology designers and developers helped to develop the four categories (with 19 items) that comprise the MARS scale: Engagement, Functionality, Aesthetics, and Information quality.

The tool was piloted on 60 mental-health related apps and showed a high degree of internal consistency and mostly high inter-rater reliability. Moderate levels of inter-rater reliability were achieved for the Ease of use and Navigation items in the Functionality category of the scale. These findings might reflect the subjective nature of these items and variability in technology literacy. It has to be noted that the item Evidence base within the Information quality scale was not evaluated in the pilot study as the item requires that the app would be trialled/tested. The authors argue that, as the impact of the apps included in the pilot study was not assessed, they excluded this item from the analysis. However, in the absence of formal evaluation data, there are measures to assess the likelihood of app effectiveness based on their content, such as inclusion of BCTs and adherence to best practice guidance (as discussed earlier).

Despite these weaknesses, the MARS scale appeared to be most suitable tool to use for the present study as a) it was the only tool that addressed an extensive variety of app design characteristic and functionalities, and b) was specifically developed to address the quality in health-related apps which was the aim of this study.

### **2.2.5 To MARS and back – experiences of applying the scientific method**

I decided that MARS would be the measure of choice for my study and I decided to pilot the scale. I watched the video training for the MARS assessment, made notes, and completed the exercise provided. I instructed my second reviewer to do the same and we used the MARS tool to assess the first app in the sample. In the next section I describe the process of using MARS and the valuable lesson I learnt from this experience.

During the discussion of the results of the pilot assessment with my co-reviewer, I realised that there was a substantial discrepancy between how we both understood the concepts addressed by the tool and how we interpreted the 5-point scale. Each of the 19 items of MARS provides a wide spectrum of app characteristics, both in the content, functionality, and appearance, most of which are subjective in nature.

I decided to attempt to make MARS as objective as possible. My first step was to contact the authors of the scale. They were very responsive but informed me that, beyond the video training there was no other information on MARS assessment. I decided it was important to devise a manual explaining the meaning of the 19 concepts with their accompanying the 5-point rating scale that constituted the MARS tool. I followed these steps:

- I transcribed the MARS training video in an attempt to make the explanation of the assessment tool more explicit
- I searched the scientific databases on various concepts addressed in the scale, such as *entertainment, interactivity, gamification, customisation (including what*

is the basic level of customising), etc. in the context of apps and, more broadly, in digital technology

- I asked other researchers that used the scale on their approach in dealing with the ambiguity of the tool but they did not provide me any additional information
- I contacted researchers in the field of health apps researchers who might assist me with creating definitions of the items
- I met with an app developer to clarify the definitions of the concepts but I was unable to get answers that would enable me to make the MARS assessment more objective

I would like to illustrate the issue with MARS using an example. The first item in the MARS tool assesses *Entertainment* of the app (Table 5):

Table 5: Entertainment item of the MARS scale

Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g. through gamification)?	
1	Dull, not fun or entertaining at all
2	Mostly boring
3	OK, fun enough to entertain user for a brief time (< 5 minutes)
4	Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
5	Highly entertaining and fun, would stimulate repeat use

*Entertaining* is a highly subjective adjective synonymous with *delightful*, *amusing*, *fascinating*, *chucklesome*, amongst others [158]. To help with my definition of the concept, I searched the area of entertainment but my research was futile as I found there is lack of research on what entertaining might mean in apps above the example of gamification.

I decided to contact a researcher who worked on defining engagement in apps and asked her about entertainment in apps. She also proposed that gamification, i.e., earning virtual rewards (points and badges) could be entertaining for the users.

Consequently, I put together a definition of gamification based on the literature available: Gamification is defined as “the use of game design elements in non-game contexts”[159]. The six core components of gamification were taken from the review of gamification elements used in health apps. These were: a) leaderboards, b) levels of achievement or rank, c) digital rewards (e.g., points, medals, badges), d) real world prizes, e) competitions/challenges, f) social or peer pressure [160].

I was momentarily satisfied with my attempts until I tried to define what 1-5 scale means in the entertainment item, and what is the difference between 4 and 5 or 1-2. This issue was not limited to the *Entertainment* item but emerged for many other definitions of the MARS tool, such as *Interest*, *Interactivity*, *Visual appeal*, etc. My determination and fixation with my ideal measure resulted in spending a substantial amount of time attempting to “transform” subjective properties to an objective assessment. Finally, I realised this was not efficient use of my time.

I stepped back, reflected and looked again at different measures to assess *the most positive experience possible* and I went back to usability assessment as I decided it was more objective than MARS. However, most of the usability assessment tools I found are complex and need to be conducted by usability experts and/or with users.

Then I reconsidered System Usability Scale (SUS), a measure I had previously dismissed almost immediately as suboptimal. It is often called “quick and dirty” assessment of usability and I thought that neither quick nor dirty were the descriptors I would like to associate with my PhD.



SUS was developed in 1986 within the industrial context to measure how people perceived the usability of the computers they worked on [161]. Although there have been many other measures of usability developed since then, the SUS is still widely used, with almost 4,000 references in various journal articles. The SUS is considered an industry standard; it is referenced in the leading user experience practice and guidelines online resource which was set up by the Digital Communications Division in the U.S. Department of Health and Human Services ([usability.gov](http://usability.gov)).

The appeal of the SUS lies in its simplicity, flexibility, and cost-effectiveness. This scale provides a single overall reference score for users' views of product's usability which is understood by people from a range of professions. As a result of this wide applicability, the SUS has been used to evaluate many types of products and services. Whilst usability testing performed in a laboratory setting is expensive [162] and time-consuming, SUS provides a low-cost alternative to solicit feedback from actual users. In the context of app reviews, SUS was used to construct an assessment for health apps in general [163] and to evaluate apps in weight loss [137].

Although developed as a tool for assessment and improvement of commercial products and services, SUS has been utilised for scientific purposes, and the tool was empirically evaluated. Studies on the psychometric properties of the SUS reported an inter-rater agreement of 0.85 [164]. In a large scale study, Bangor, Kortum [165] analysed 10 years' worth of SUS data, including 206 studies which represented 2,324 user surveys. They found high reliability of the instrument (Cronbach's alpha of 0.91).

Attempts were made, using factor analysis, to identify the distinct concepts measured by the SUS [165, 166]. The number of unique concepts, also called the dimensionality of the scale, is somewhat controversial; some researchers reported only one significant

factor for the 10 items that comprise the SUS scale suggesting that the tool measures one dimension: Usability [165]. However, Lewis [166] argued that SUS measures, in fact, two dimensions: Usability and Learnability. For the purpose of this study, the SUS will be analysed as a single scale reflecting usability, as originally intended by Brooke [162] .

As the use of the SUS increased, the interpretation of the total score (and average total score) became standardized. The scale ranges from 0 to 100, so the mean score has the same range. A common interpretation emerged that is similar to a university grading system, for example, score of 90 to 100 was an A, 80-89 was a B, etc. Although these thresholds were practical and widely used, they were not validated or anchored to concepts such as user preference. Bangor, Kortum [165] attempted to address the issue and conducted a study where an additional descriptive item was added to the SUS scale. This new item comprised of seven ordered adjective phrases (from best imaginable to worst imaginable). The SUS tool including this additional item was administered to 964 participants. They found that the adjective rating scale was strongly related to the SUS scores (correlation of 0.806), which would indicate that the thresholds are appropriate. This study will use these descriptors to interpret the app usability scores.

The SUS scale's simplicity is a great advantage but it is also restrictive in its low granularity of the data obtained. I realised that I need to make the trade-off of what is ideal and what is possible in the context of this review. Based on what I have learnt, I argue that in researcher-led review of apps, the most objective measure of experience possible should be used to assess the quality of apps in order to yield meaningful results. More subjective measures pertaining to user experience should be assessed with the users themselves.

---

End of the reflective log

---

### **2.2.6 Previous reviews and content analyses of quality in PA apps**

The most recent review and content analysis of apps aimed at increasing PA described the frequency of BCTs used in the sample of 64 apps [145]. They found that, on average five BCTs were used in each app. The study excluded apps which did not provide tailored feedback and those that did not follow the recommended PA levels. The study also used the previous version of BCT Taxonomy which included 26 items [144].

Similarly to Middelweerd et al.'s study [145], Conroy, Yang [167] rated the presence of BCTs in 167 PA apps' descriptions and found that each app incorporated fewer than four BCTs. The most common BCTs included instruction, modelling, providing feedback on performance, goal-setting, and planning social support/change. This study used the limited 40-item CALOR-RE Taxonomy [168], and apps were coded based on their description, and not the downloaded content.

Yang, Maher [139] assessed the presence of BCTs in PA apps using the most recent BCT Taxonomy (v1; Michie et al., 2013) and tested for difference between paid and free apps. They found that, on average, only 6.6 of 93 BCTs in the Taxonomy were present in each app. They found no difference between paid and free apps.

Cowan, Van Wagenen [169] assessed the inclusion of various health behaviour theory constructs in PA apps and found that only a small number of apps included behaviour theory constructs. In addition, the study suggested higher quality apps were more expensive. The assessment was limited to the four major health behaviour theories, and only apps available in iTunes store were evaluated. In addition, the study excluded apps only with overall  $\leq 4$  star rating, and those that were not available on iPad.

Knight, Stuckey [170] assessed adherence to PA guidelines in PA apps and presence of selected technological features in the sample and found that none of the apps implemented the PA guidelines on aerobic PA, and seven out of 379 adhered to the guidelines for resistance training. The assessment included a random selection of apps from the two pre-specified groups: endorsed by a non-commercial agency and only those that have ratings of  $\geq 4.5$ .

In conclusion, previous studies described the presence and frequency of BCTs in PA apps [139, 145, 167], inclusion of health behaviour theory constructs [169] and adherence of app content to PA guidelines [170]. Of the studies that used a systematic classification of the content of behaviour change interventions, only one is the latest version of the taxonomy (BCT Taxonomy v1; Michie et al., 2013). All the previous reviews of PA apps focused on the behavioural content as operationalised by inclusion of theory or behaviour change techniques. Inclusion of theory is an important aspect of quality assessment. However, as discussed, various different quality indicators should be considered when assessing app quality, such as design quality and data security.

The novelty of this study lies in the systematic adoption of healthcare quality indicators to assess the current publicly available PA apps. As such, this study is looking at the potential efficacy of those apps, their data safety and privacy, and the user experience of those apps. At the time of writing this chapter, no other studies have looked at such comprehensive definition of quality of apps.

### **2.2.7 Aims and objectives**

The overall aim of this study was to ascertain the quality of current PA apps available on the market using the healthcare quality criteria of a) safety, b) effectiveness, and c) provision of the most positive experience possible. Table 6 summarises the healthcare quality indicators used in this study.

The specific objective of the study were:

- To identify the most popular PA apps on the market
  - To assess the privacy and security of the users' data
  - To ascertain the likely effectiveness by quantifying the presence of BCTs in the apps
  - To describe the quality of app development and the evaluation process in terms of organisational affiliation, expert and user involvement, and the evidence of evaluation in peer-reviewed journals
  - To describe the user experience by assessing the user ratings and the usability of the apps
- To assess if the free apps differ from the paid apps

Table 6: The summary of the healthcare quality indicators as applied to PA apps

Quality indicator of healthcare	Applying the indicator to health apps
Safety	Privacy and security of data
Effectiveness	Behaviour Change Techniques [116] Development and evaluation process: <ul style="list-style-type: none"> <li>– Organisational affiliation</li> <li>– Expert involvement</li> <li>– User involvement</li> <li>– Evidence of scientific evaluation</li> </ul>
Positive experience	User ratings System Usability Scale [162]

## 2.3 Methods

### 2.3.1 Study design

This study was a review and a content analysis of the most popular PA apps on the market. The *Quality and Risk of Bias Checklist for Studies That Review Smartphone*

*Applications* was used to ensure that methods for apps' review were adequately described [83].

### **2.3.2 Sample identification**

The sample of apps for increasing PA was obtained from the UK's versions of the iTunes and Google Play (GP) stores on 17/10/16. As there are mixed findings on the association between price and inclusion of behaviour change components [139, 169, 171] both free and paid apps were included in the study. Popularity of apps was defined as being in the top 100 ranking (paid and free) in the "Health & Fitness category" in both stores. This method of assessing popularity has been used in other studies assessing apps [137, 172, 173], and it was selected in order to simulate the user experience of browsing the store to select a health app. The first 100 apps were considered as users are more likely to focus on the top results and rarely examine the search results thoroughly [174]. In total, 400 apps' titles and descriptions (100 iTunes free + 100 iTunes paid + 100 GP free + 100 GP paid) were screened against the eligibility criteria (see Appendix A for the 400 apps used for sample identification process). Duplicate apps (the same apps in both stores), were considered if the duplicates were within the 400 highly ranked apps. The apps were screened by two reviewers against the following eligibility criteria:

Apps were included if:

- Their main goal was to increase PA
- They were targeted at healthy adults
- They had stand-alone functionality (did not need to be linked with peripheral devices, such as pedometers; and did not require paid membership in order to access the app)

Apps were excluded if:

- The primary aim of the app was focused on multiple behaviours or behavioural outcomes (e.g., apps targeting both PA and diet in equal measures) as it would

be difficult to compare apps that target multiple health behaviours [169] with those targeting a single behaviour;

- The target population was patients with a specific health condition, as these users were likely to have different needs and motivation to healthy adults;
- They were sold as part of a pack (“bundle”) as it was not impossible to extract the popularity ranking from a pack of multiple apps. These apps are very often positioned within the ranks separately.

### **2.3.3 Sample assessment**

The apps were downloaded onto an iPhone SE and 6 (running iOS 10.2.1 and 9.3.4 software respectively) and Android Samsung Galaxy S6 and J5 (running 6.0.1 and 5.1.1 software respectively), and assessed independently by two reviewers (PB and GA) using an extraction form (Appendix B for data extraction form). Each app was left running in the background for two days for the assessors to explore any reminders/notifications. If two apps were identified as duplicates and there appeared to be consistency of design and content between both operating systems, the apps were assessed on iPhone only.

### **2.3.4 Data extraction and measurement**

For the purpose of this study, PA is defined as any aerobic activity that can produce moderate (such as brisk walking, riding a bicycle on a level ground) to vigorous activity (such as running, riding a bicycle on hills), or muscle-strengthening activity (such as working with resistance bands, yoga) [89, 175].

#### **2.3.4.1 Descriptive data**

The following descriptive data were extracted from both app stores: app’s name, brief description, type of PA targeted (e.g., running, walking, whole body workout), platform on which the app was available, developer’s name, rank, number of ratings, cost, size, last update, and version. As there is little research looking at what influences user ratings

in the app stores, all the potential variables that might affect user experience were extracted from the app stores.

#### 2.3.4.2 Data privacy and security assessment

The privacy and security assessment (Table 7) was devised for this study and was based on the recommendations of ICO and OTA. It comprises of eight questions evaluating the availability, accessibility of privacy policy, data gathering and sharing practices, and data security as it is discussed in the privacy statement. The manual for the data extraction of the data privacy and security assessment is available in the Appendix C.

Table 7: Data privacy and security assessment

Q1: Availability	Is there privacy information available? (only continue if answered 'Yes')	Yes/No
Q2: Availability	Is the privacy information available without the need to download the app? (example: app store, via a link to the privacy policy or the app's website)	Yes/No
Q3: Availability	Is the privacy information available within the app?	Yes/No
Q4: Accessibility	Is there a short form notice (in plain English) highlighting key data practices which are disclosed in detail in the full privacy policy?	Yes/No/NA*
Q5: Accessibility	Is the privacy policy available in any other languages?	Yes/No
Q6: Data gathering	Does the app collect Personally Identifiable Information?	Yes/No/NS
Q7: Data sharing	Does the app share users' data with 3rd party?	Yes/No/NS
Q8: Data security	Does the app say how the users' data security is ensured? e.g. encryption, authentication, firewall system	Yes/No

Notes: NA, not applicable; NS, not specified; \*NA in Q4 was indicated when the policy statement was concise



### **2.3.4.3 Behaviour Change Techniques**

Michie et al.'s 2013 BCT Taxonomy (v1) was used to assess the number of BCTs in each app and the frequency of each BCT in the app sample overall. The coding manual provides guidelines to investigate the presence of 93 BCTs in behaviour change interventions and has been used in previous studies that aimed to characterise BCTs in health apps [136-138, 167, 176]. In line with the instructions, each BCT was coded as Absent, Present ++ (BCT present beyond all reasonable doubt), Present + (BCT present in all probability but evidence unclear). The evidence for the presence of the BCTs in each app was recorded.

### **2.3.4.4 App features**

Health apps use technology-enhanced features to deliver BCTs in order to influence behaviour. The apps were categorised according to the features offered by the app, e.g., PA tracking, reminders, app community, data visualisation, etc.). The terms *feature* and *functionality* are often used interchangeably [137, 177, 178]. As this study focuses on the visible properties of an app from the user's point of view, a user-centric definition of features was utilised: "For systems with a large number of internal states, it is easier, and more natural, to modularize the specification by means of features perceived by the customer".

#### **2.3.4.4.1 App features and mapping of the features and the BCTs**

In this study, BCTs were mapped onto the features of the PA health app sample in order to gain a better understand of the features that support BCTs to increase PA. To the writer's best knowledge, there is no list of features that are commonly used in PA apps. 10 apps that were extracted for the power size calculation (required for Study 2 and discussed in Chapter 3, section 3.3.6: Sample size justification) were used to compile a list of features. The list was continuously updated throughout the app extraction process in order to accommodate for new features that were found in the PA app sample.

#### **2.3.4.5 Quality of development process and evidence for evaluation**

The following characteristics of the app content development were extracted: organisational affiliation (university, medical, government or other non-profit institutions); health or behaviour change expert involvement (e.g., health professional, behaviour change specialist, etc.); evidence for user involvement in the development of an app; evidence of evaluation assessed by presence of any peer-reviewed studies assessing the app. Studies providing data on any scientific evaluation of the apps included in the sample were identified by searching for the name of the app using the following scientific databases: PubMed, ACM Digital Library, IEEE Xplore, Google Scholar.

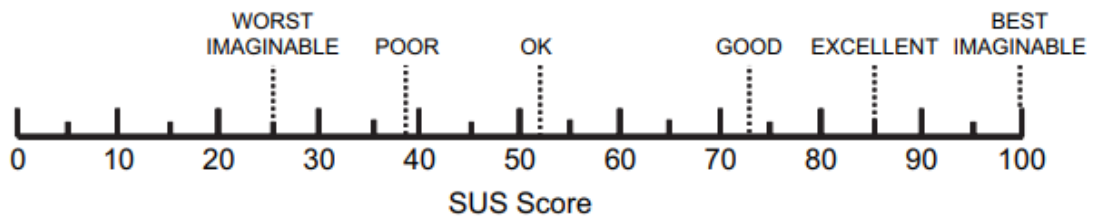
#### **2.3.4.6 User Ratings**

In app stores, the user can award 1 to 5 stars to the app. The average star rating was calculated by summing the number of stars (standard range from 1 to 5 stars) and dividing them by the number of users who submitted ratings.

#### **2.3.4.7 SUS**

SUS consists of 10 items that are ranked on a 5-point Likert scale, from “strongly disagree” to “strongly agree”. Half of the statements are negatively loaded. In this study, I made 2 changes to the original scale. First, the wording of the 8th statement was changed from "cumbersome" to "awkward" as recommended [165, 166, 179]. Second, the word "system" was replaced by "app" to make the scale applicable to the sample in this study. The interpretation of the SUS score used the thresholds proposed and validated by Bangor, Kortum [165] (Figure 5).

Figure 5: Descriptive ratings representing the SUS score (reproduced with permission from [165])



When assessing the usability of the apps, the reviewers followed the Nielsen’s five quality components which define usability [180] which was marginally adapted to reflect the review process instead of assessing the apps with users (Table 8). The reviewers kept a descriptive record of their rationale of each SUS score which consisted of notes on overall usability and any issues they experienced with app usability (Appendix D displays SUS score descriptive record). The SUS score for each app from the 2 reviewers was averaged.

Table 8: Five questions used in the review when assessing usability

<b>Q1: Learnability</b>	How easy is it for the reviewer to accomplish basic tasks the first time they encounter the app?
<b>Q2: Efficiency</b>	Once the reviewer has learned the app, how quickly can they perform tasks?
<b>Q3: Memorability</b>	When the reviewer returns to the app after a period of not using it, how easily can they re-establish proficiency?
<b>Q4: Errors</b>	How many errors do the reviewers make, and how easily can they recover from the errors? (In the case of this assessment, the reviewers will focus on crashes/bugs/broken features, etc.)
<b>Q5: Satisfaction</b>	How pleasant is it to use the app?

### **2.3.5 Inter-rater reliability assessment**

Discrepancies between the reviewers were resolved through discussion and consultations with an expert in behaviour change (LA). The a priori strategy for assessing the coding agreement was to complete the extraction of data for 10 apps, compare the results and discuss any discrepancies in understanding of the measures before extracting the rest of data. Following the discussions about the BCT extraction, it was decided that a BCT Manual will be devised to facilitate the extraction of BCTs from PA apps. The Manual was based on the experiences of extracting the BCTs from the first 10 PA apps and was refined after the discussions of discrepancies between the 2 reviewers. The BCT Manual is presented in Appendix E.

Inter-rater reliability was assessed for the presence (either ++ or +)/absence of the BCTs by calculating Cohen's kappa statistic [181] for each item. In addition, 'Prevalence and bias adjusted kappa' (PABAK) [182] was assessed for the presence/absence of BCTs. The occurrence of high prevalence of negative agreement (when both rates agree that the BCT is absent) is very likely in the context of inclusion of BCTs in an app. When high prevalence of the identical response is seen, the kappa value results in low proportion of agreement, although the observed agreement is high [183] .

Figure 6 below shows the comparison of kappa and PABAK agreement measures. For example, in the 3<sup>rd</sup> instance the proportion of observed agreement that the BCT is absent is 0.83 (5/6), but the kappa shows 0 agreement. PABAK adjusts for the high prevalence of responses showing substantial agreement. PABAK has been used for psychometric properties of BCT Taxonomy and other studies concerned with app assessment [138, 184].

Figure 6: Hypothetical examples of coders' agreement of 6 BCTs with a comparison of kappa and PABAK.

Agree about 3/6				Agree about 4/6				Agree about 5/6					
Coder2	Coder1			Total	Coder2	Coder1			Total	Coder2	Coder1		
	Present	0	2			Present	4	2			Present	0	0
	Absent	1	3			Absent	0	0			Absent	1	5
	Total	1	5			Total	4	2			Total	1	5
	% agree	50%				% agree	67%				% agree	83%	
$\kappa$	-0.29	None	$\kappa$	0.00	None	$\kappa$	0.00	None					
PABAK	0.00	None	PABAK	0.33	Fair	PABAK	0.67	Substantial					
Coder2	Coder1			Total	Coder2	Coder1			Total	Coder2	Coder1		
	Present	2	0			Present	4	1			Present	1	0
	Absent	3	1			Absent	1	0			Absent	1	4
	Total	5	1			Total	5	1			Total	2	4
	% agree	50%				% agree	67%				% agree	83%	
$\kappa$	0.18	Slight	$\kappa$	-0.20	None	$\kappa$	0.57	Moderate					
PABAK	0.00	None	PABAK	0.33	Fair	PABAK	0.67	Substantial					

### 2.3.6 Data analysis

The number of BCTs in the apps was summarized using the mean, standard deviation, median, 25<sup>th</sup> and 75<sup>th</sup> percentiles, and the maximum and minimum. Similar statistics were used to summarize user ratings, cost, size, and SUS score. Proportions were used to summarize the variables: data privacy and security, organisation affiliation, expert and user involvement and the evidence of evaluation in peer-reviewed journals.

The summary descriptive tables were presented for each store for free and paid apps separately and in total, as app stores have separate rankings based on the cost. Free and paid apps were compared as there is some evidence indicating a relationship between price and other app characteristics, such as the inclusion of BCTs [139, 169, 171]. Specifically, to assess if there was a differences in store characteristics between free and paid apps, t-test was used to compare the average user ratings, size and the number of BCTs; Wilcoxon test was used to compare the number of ratings; Fisher's

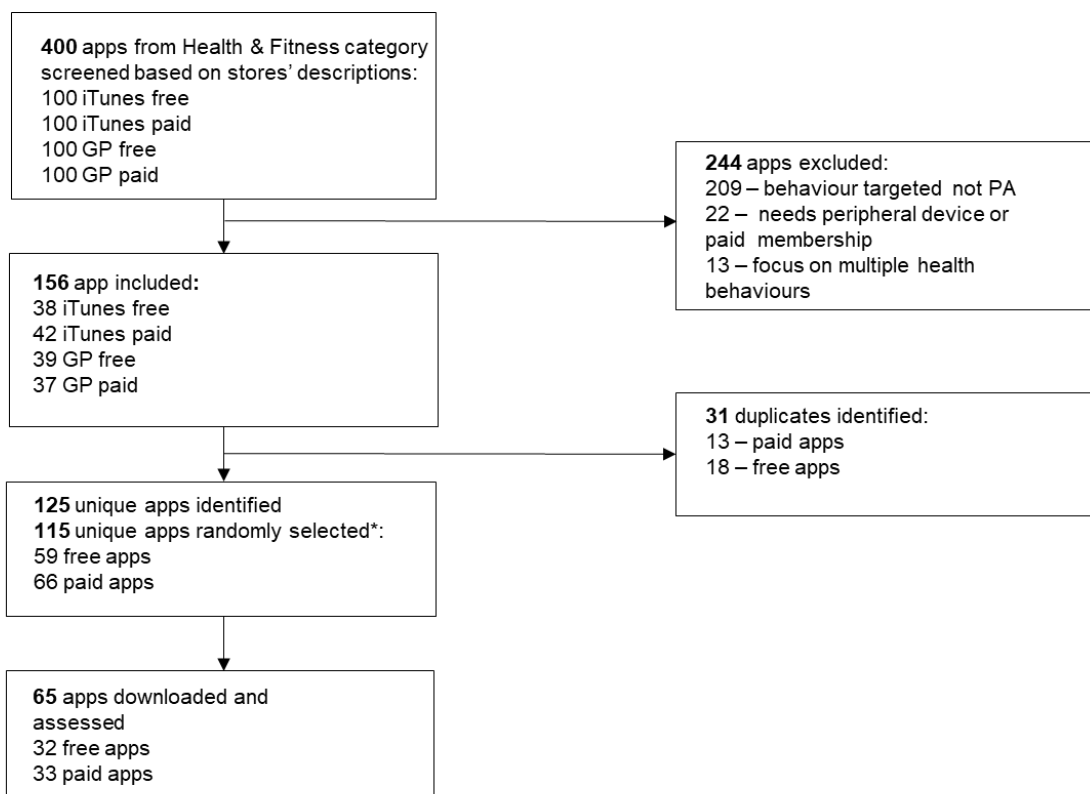
exact was used for last update (<3 months; 3-6 months; >6m months), organisational affiliation, expert and user involvement, and presence of any peer-reviewed studies.

## **2.4 Results**

### **2.4.1 *Sample identification results***

Out of 400 apps, 244 apps were excluded (209 apps did not target PA, 22 apps needed a peripheral device or paid membership in order to use the app and 13 apps focused on multiple health behaviours), and 156 met the inclusion criteria. Thirty one duplicates were found yielding a sample of 125 unique apps. From this sample, 10 apps were used to conduct the sample size calculation which was a necessary step for Study 2 which assessed an association between quality indicators and user ratings (described in detail in Chapter 3, section 3.3.6: Sample size justification). Consequently, a sample of 65 apps was required as indicated by the sample size calculations. 32 free and 33 paid apps, were randomly selected using the random number generator function in Excel. A flow chart of the app selection process is summarised in Figure 7.

Figure 7: Flow chart of the apps included in the analysis



Notes: GP, Google Play; \* 10 apps were assessed prior to conduct sample size calculations, hence 115 were randomly selected for the sample

## 2.4.2 Sample assessment results

### 2.4.2.1 Descriptive data

Individual level data for each app is presented in Appendix F. The descriptive data (user ratings, number of ratings, cost, size, and last update) are presented in Table 9 for iTunes and Table 10 for GP. There were no statistically significant differences in the number of ratings, cost, size, and last update between the free and paid apps in either iTunes or GP store.

#### Number of ratings

In iTunes, there was a wide range of the total number of ratings given to the apps, from 11 to 24,530 ratings. Free apps were rated more than paid apps (median of 758 for free

versus 127 median number of ratings for paid apps). The total number of ratings given to each app varied substantially in GP, ranging from 7 to 625,077. Similar to iTunes, free apps were rated more than paid apps (median number of 44,923 for free versus 1,720.5 median number of ratings for paid apps). There was a substantial difference in the number of ratings between iTunes and GP store with median number of ratings of 550 in iTunes and 5856 in GP. 50% of the apps had between 86 and 1,719 ratings in iTunes, while in GP 50% of apps had between 1,475 and 78,204.

### **Price**

The average price of paid iTunes apps was £2.50. Apps in GP were slightly more expensive than in iTunes, with an average price of £3.60.

### **Size**

In iTunes, the size of the apps varied substantially with the mean of 91.8 and SD of 64.1 MB. Paid apps were slightly larger than free apps (the mean of 94.9 in comparison to 88.4). The paid apps in GP were also larger than the free apps (the mean of 43.4 in comparison to 28.4). The total mean MB in GP was 43.4 and SD of 28.2, and there was a difference in the sizes of apps between the 2 stores, with iTunes apps being substantially larger (mean of 91.8 in iTunes versus 43.4 in GP).

### **Last update**

More than half of the free apps (61.9%) were updated within the last three months in comparison to 29.2% of paid apps in iTunes. Within the paid group, 41.7% were updated more than 6 months ago. Similarly, in GP free apps, the largest group (76.2%) were updated in the last 3 months. Within the paid apps, 43.8% were updated within the last 3 months, 18.8% between 3-6 months, and 37.5% of apps were updated > 6 months ago.



Table 9: Descriptive data – iTunes

		Free - iTunes (N=21)	Paid - iTunes (N=24)	Total - iTunes (N=45)	p- value
Avg. user rating (1-5 stars)	Mean ± SD	4.1 ± 0.8	4.3 ± 0.6	4.2 ± 0.7	0.251
	Median	4.4	4.6	4.4	
	25, 75%tile	4.0, 4.6	4.0, 4.8	4.0, 4.6	
	Min, Max	2, 5	3, 5	2, 5	
Number of ratings	Mean ± SD	3408.4 ± 5848.4	773.7 ± 1187.0	2031.2 ± 4289.7	0.158
	Median	758	127	550	
	25, 75%tile	438.0, 3698.0	47.0, 1247.0	85.5, 1719.0	
	Min, Max	14, 24530	11, 3845	11, 24530	
Cost - iTunes	Mean ± SD		2.5 ± 1.5		
	Median		2.3		
	25, 75%tile	N/A	1.5, 3.0	N/A	
	Min, Max		1, 8		
Size (MB)	Mean ± SD	88.4 ± 49.8	94.9 ± 75.4	91.8 ± 64.1	0.739
	Median	74.3	83.3	82.2	
	25, 75%tile	52.0, 131.0	61.7, 102.0	58.1, 104.0	
	Min, Max	11, 164	9, 376	9, 376	
Last Update	< 3 mo	13 (61.9%)	7 (29.2%)	20 (44.4%)	0.084
	3-6 mo	3 (14.3%)	7 (29.2%)	10 (22.2%)	
	> 6 mo	5 (23.8%)	10 (41.7%)	15 (33.3%)	

Notes: MB, megabytes; N/A, not applicable; mo, months

Table 10: Descriptive data – GP

		<b>Free - GP (N=21)</b>	<b>Paid - GP (N=16)</b>	<b>Total - GP (N=37)</b>	<b>p- value</b>
Avg. user rating (1-5 stars)	Mean ± SD	4.4 ± 0.5	4.4 ± 0.3	4.4 ± 0.4	0.784
	Median	4.5	4.5	4.5	
	25, 75%tile	4.4, 4.6	4.4, 4.6	4.4, 4.6	
	Min, Max	2, 5	4, 5	2, 5	
Number of ratings	Mean ± SD	119000.7 ± 165085.0	14457.9 ± 43700.8	73793.0 ± 136723.2	1
	Median	44923	1720.5	5856	
	25, 75%tile	5827.0, 199596.0	384.5, 6452.0	1475.0, 78204.0	
	Min, Max	206, 625077	7, 177277	7, 625077	
Cost - Google Play	Mean ± SD		3.6 ± 2.3		
	Median		2.7		
	25, 75%tile	N/A	2.3, 5.0	N/A	
	Min, Max		1, 9		
Size (MB)	Mean ± SD	28.4 ± 21.2	43.4 ± 34.2	34.9 ± 28.2	0.109
	Median	26.8	31.5	29.6	
	25, 75%tile	12.2, 38.5	27.7, 54.0	15.4, 43.9	
	Min, Max	2, 73	1, 145	1, 145	
Last Update	< 3 mo	16 (76.2%)	7 (43.8%)	23 (62.2%)	0.120
	3-6 mo	1 (4.8%)	3 (18.8%)	4 (10.8%)	
	> 6 mo	4 (19.0%)	6 (37.5%)	10 (27.0%)	

Notes: MB, megabytes; N/A, not applicable; mo, months

### **The highest-ranked apps**

The top-ranked free app in both iTunes and GP, was Fitbit by Fitbit, Inc. and the highest ranked paid app was *7 Minute Workout Challenge* by *Fitness Guide Inc.*, also in both stores.

#### **2.4.2.2 Type of PA apps**

In order to characterise the sample, the apps were categorised into five groups that were similar in their primary function: workouts; movement trackers; running programmes; pedometer-based apps; and interval timers.

##### **Workouts**

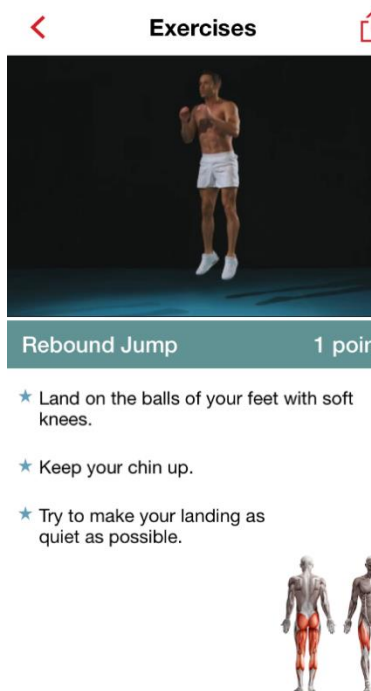
Of the 65 apps assessed, the most common type of apps was prescribed workout apps (N= 31). These apps had specific exercises that were often grouped into workouts. The primary function of the apps in this group was to track the completion of the workout based on timers or user-logged PA information. The distinctive features of the workout apps are videos and audio instructions; timers to guide the user when to exercise and rest; and summary reports on the frequency of workouts completed.

Within this category of app, 20 apps included various workouts using either equipment or bodyweight; 5 apps tracked one type of workout, the *7 Minute Workout*, developed by the researchers from the Human Performance Institute [185] who combined aerobic and resistance training into short high intensity interval training (HIIT); 4 apps were focused on abdominal muscles only; one app focused on the goal of achieving 100 push-ups (Figure 8).

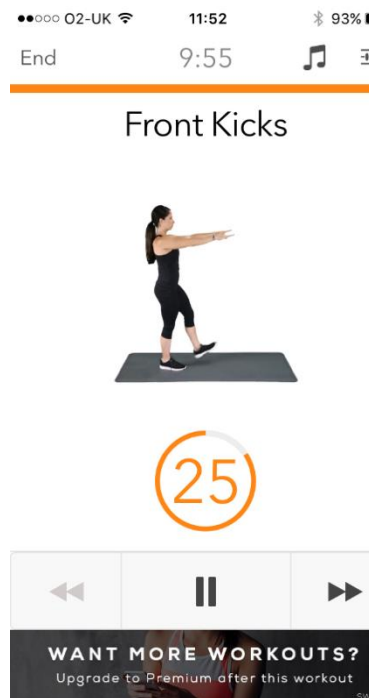
Figure 8: Examples of the features in the workout apps.

1. Video instructions in Adrian James High Intensity Interval Training by Adrian James Nutrition Ltd;
2. Timers to guide the user when to exercise and rest in Sworkit by Nexercise ;
3. Summary reports in 7 Minute Workout Challenge by Fitness Guide Inc.

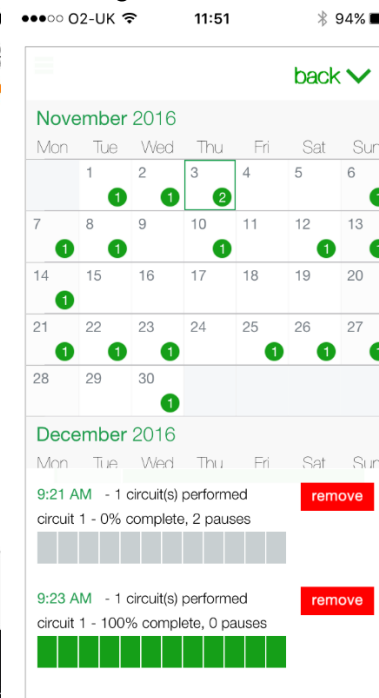
### 1. Adrian James High Intensity Interval Training



### 2. Sworkit



### 3. 7 Minute Workout Challenge



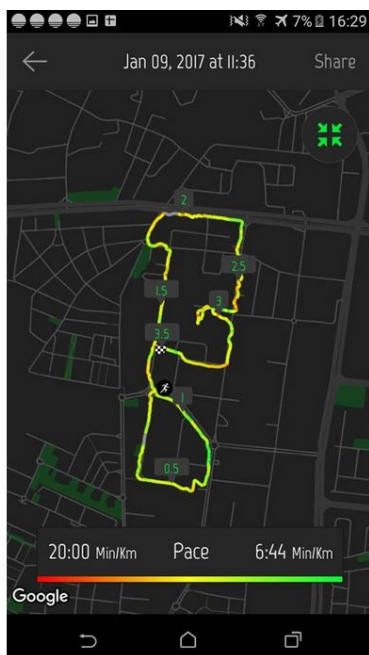
## Tracking of movement

The second category was tracking of movement (N= 13). The primary function of the app in this category is to automatically track PA, primarily running but also walking or cycling. The distinctive features of these PA tracking apps are mapping of the running/walking/cycling route; visualisations of the data, such as charts on pace, distance, duration, elevation, etc.; and access to a database of running routes (Figure 9).

Figure 9: Examples of the features in the tracking of running apps

1. Mapping of the running route in *Running Distance Tracker +* by *Fitness22 Ltd*;
2. Data visualisations of the run in *Map My Run* by *Under Armour*
3. Database of running routes - *WalkJogRun* by *Almost Awesome Inc.*

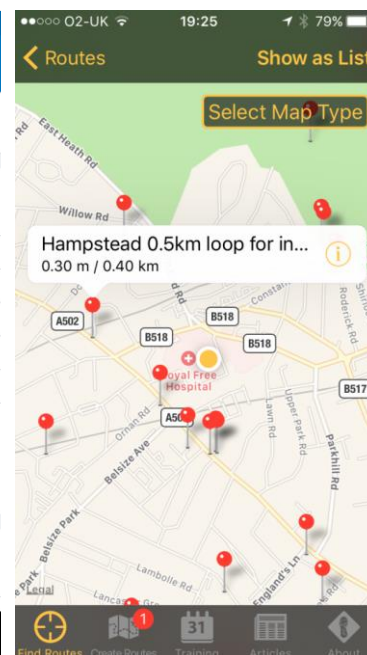
1. Running Distance Tracker +



2. Map My Run



3. WalkJogRun



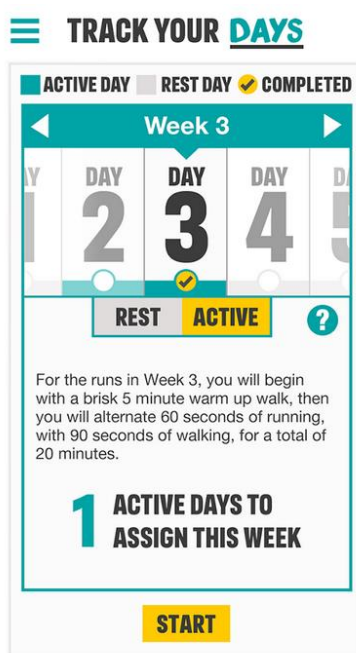
## Running programmes

The third group of apps was running programmes (N= 12). The primary function of these apps was to reach a pre-specified running goal. The goal was achieved by incremental increase in run to walk ratio, for example 5 km running programmes set a goal for training 30 minutes/3 times a week for 9 weeks. By the end of the 9<sup>th</sup> week the user is expected to run for 30 min without pausing. The distinctive features of the running programme apps were scheduling of time-specific goal-oriented running sessions; audio-guidance during running; and reports on sessions completed (Figure 10).

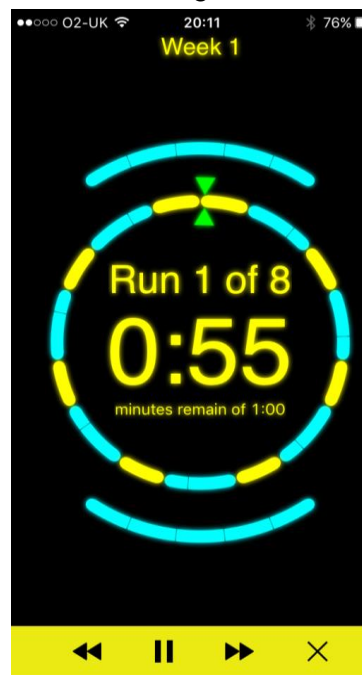
Figure 10: Examples of the features in the Running programmes

1. Scheduling of set time-specific, goal-oriented running sessions in *One You Couch to 5K* by PHE;
2. Audio-guided running sessions on when to walk/run in *Get Running (Coach to 5K)* by Benjohn Barnes;
3. Reports on sessions completed in *Couch to 5k Runner* by Fitness22 Ltd

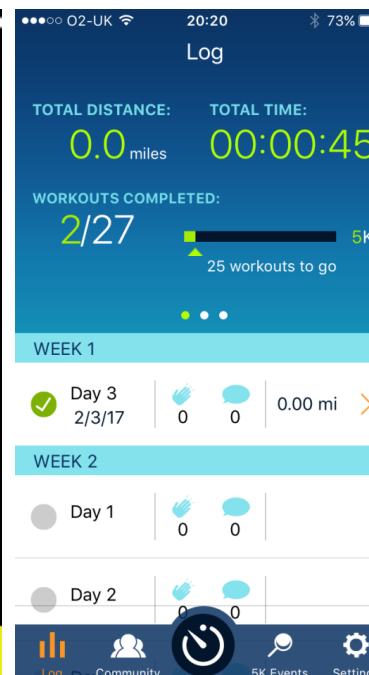
1. One You Couch to 5K



2. Get Running



3. Couch to 5k Runner

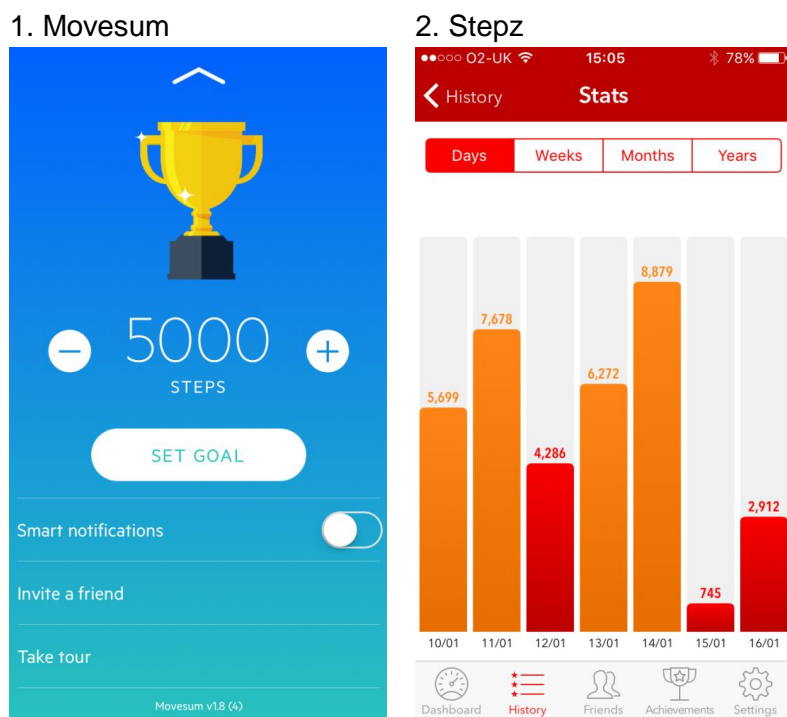


### Pedometers-based apps

The fourth distinctive app category was pedometer-based apps (N= 6). The primary function of the app in this group was to provide the user with the feedback on their step count. The distinctive features of these app were setting step goals; provision of feedback in the form of summary reports and visualisations (daily/weekly etc.) of the steps recorded (Figure 11).

Figure 11: Examples of the features in the pedometer-based apps group

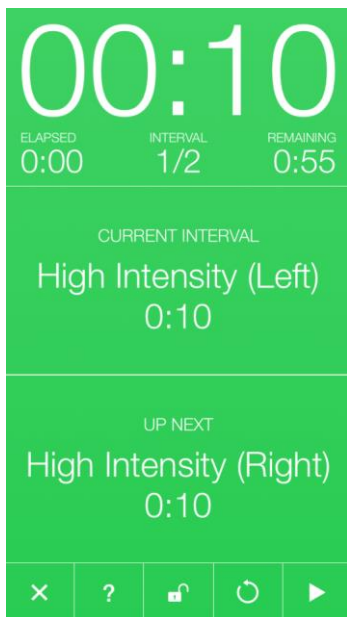
1. Setting step goals in *Movesum* by *Lifesum AB*; provision of feedback and summary reports on steps recorded in *Stepz* by *VisualHype GmbH*.



### Interval timers

The fifth category was interval timers (N= 3). The primary function of these apps was to track the work and rest periods during workout sessions. The app does not provide PA instruction but enables the user to set their own interval workout routine.

Figure 12: Example interval timer app in Interval Timer by *Deltaworks*



#### 2.4.2.3 Data privacy and security

The number of apps that had privacy policy is presented in Table 11. Within those apps that the privacy policy was available, the results of the data privacy and security assessment are summarised in Table12.



Table 11: Descriptive data for the availability of the privacy policy (N=65)

		<b>Free (N=32)</b>	<b>Paid (N=33)</b>	<b>Total (N=65)</b>
Does the app have a privacy policy?	No	8 (25%)	11 (33.3%)	19 (29.2%)
	Yes	24 (75%)	22 (66.6%)	46 (70.8%)

Table 12: Descriptive data for the data privacy and security assessment (N= 46)

		<b>Free (N=24)</b>	<b>Paid (N=22)</b>	<b>Total (N=46)</b>
Is the privacy information available without the need to download the app?	Yes	24 (100.0%)	22 (100.0%)	46 (100.0%)
Is the privacy information available within the app?	No	13 (44.8%)	16 (55.2%)	29 (63.4%)
	Yes	11 (64.7%)	6 (35.3%)	17 (36.6%)
Is there a short form notice (in plain English) highlighting key data practices?	N/A	3 (12.5%)	6 (27.3%)	9 (19.6%)
	No	17 (70.8%)	16 (72.7%)	33 (71.7%)
	Yes	4 (16.7%)	0 (0.0%)	4 (8.7%)
Is the privacy policy available in any other languages?	No	20 (83.3%)	21 (95.5%)	41 (89.1%)
	Yes	4 (16.7%)	1 (4.5%)	5 (10.9%)
Does the app collect Personally Identifiable Information?	N/S	1 (4.2%)	0 (0.0%)	1 (2.2%)
	No	2 (8.3%)	6 (27.3%)	8 (17.4%)
	Yes	21 (87.5%)	16 (72.7%)	37 (80.4%)

Does the app share users' data with a 3rd party?	N/S	1 (4.3%)	1 (4.5%)	2 (4.4%)
	No	2 (8.7%)	8 (36.4%)	10 (22.2%)
	Yes	20 (87.0%)	13 (59.1%)	34 (73.9%)
Does the app say how the users' data security is ensured? e.g. encryption, authentication, firewall, etc.	No	13 (54.2%)	14 (63.6%)	27 (58.7%)
	Yes	11 (45.8%)	8 (36.4%)	19 (41.3%)

Notes: N/A, not applicable; N/S, not specified

### Availability

The privacy policy was available for 46 (70.8%) out of 65 apps overall. In one case, the link to the privacy policy was provided but did not work, and the app was indicated as not having a privacy policy. For those apps that had a policy, 46 could be accessed without the need to download the app. In 17 instances, the privacy policy could be accessed within the app as well.

### Accessibility

Only 4 (8.7%) apps had a short form privacy and security notice that highlighted key data practices which were disclosed in details in the full privacy policy. There were 9 instances where the short form notice was not applicable due to the policy already being concise. Within these there were only 2 instances of policy that was simple and written in plain English, the developers were Fitness22 and Adrian James Nutrition Ltd., e.g.:

No personal details are passed on to third parties nor shared with companies / people outside of the company that operates this website.

[Adrian James Nutrition Ltd]

Multi-lingual policies were rare with only 5 apps having a policy in another language. Apps that were developed outside the UK were more likely to provide multi-lingual policies.

### **Data gathering and sharing**

Most of the apps (73.3%) reported collecting Personally Identifiable Information. In one instance the developer did not discuss the data gathering practices. There were rare instances where the policy clearly stated that no data is being collected by the app, for example:

What information does Fitness22 collect? Nothing. None. Nada.

[Fitness22]

In 33 instances the developers stated that they share the data they gather with 3<sup>rd</sup> parties. There were two instances where the developer did not discuss data sharing practices. In many cases the policies stated that “data shall not be shared, except for” followed by a list of exceptions that were vague and general. In these instances, the reviewers considered that the data were shared by the 3<sup>rd</sup> party, for example:

Lifesum will not share your personal data with third parties without your permission, except in the limited circumstances provided below. Personal data collected from you may be shared with our affiliates, agents and business partners. We may disclose your personal data in order to comply with a legal or regulatory obligation, if we reasonably believe that this is required by law, regulation or other legislation, or in order to protect and defend Lifesum, our business partners or users' rights and interests. [Lifesum]

## **Data security**

Only 41.3% of the apps described how the users' data would be protected. Most of the app developers stated that data safety is important to them but did not state how the data security is ensured, for example:

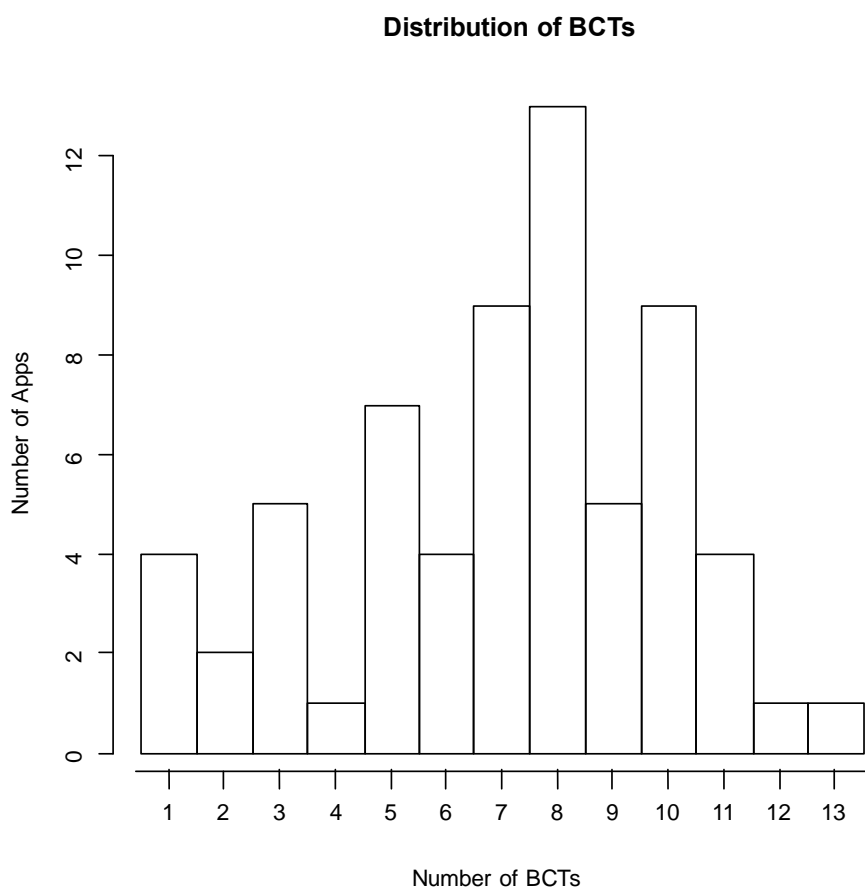
We endeavour to use and keep under review appropriate technical and organizational measures (including staff training and awareness) in order to protect against unauthorised or unlawful processing of your personal data, including unauthorised destruction, alteration or disclosure or access. We shall only hold personal data for as long as is necessary or as required by law. [3d4medical]

### **2.4.2.4 The presence of BCTs**

#### **Inter-rater reliability**

As described in the methods, the first 10 apps that were used to identify any inconsistencies in understanding of the BCTs were excluded from the agreement assessment. Consequently, inter-rater reliability was assessed on 55 apps. There was 'almost perfect' agreement: (PABAK) =0.94, 95% CI 0.93 to 0.95, kappa =0.78 ('substantial'), 95% CI 0.75 to 0.81. The distribution of the BCTs was fairly symmetric with no outliers (Figure 13).

Figure 13: the distribution for BCTs combined for free and paid apps.



The total number of BCTs for free and paid apps sample was similar. The median number of BCTs was 7 for free and 8 for paid apps. Every app contained at least 1 BCT, and the maximum number of BCTs was 12 for free and 13 for paid apps. The Table 13 presents the descriptive statistics for BCTs.

Table 13: descriptive statistics for the inclusion of the BCTs

		<b>Free (N=32)</b>	<b>Paid (N=33)</b>	<b>Total (N=65)</b>	<b>P-value</b>
Total BCTs	Mean ± SD	6.6 ± 3.0	7.5 ± 2.9	7.0 ± 2.9	0.209
	Median	7	8	8	
	25, 75%tile	5.0, 8.0	6.0, 10.0	5.0, 9.0	
	Min, Max	1, 12	1, 13	1, 13	

Figure 14 shows the frequency of the common BCT groups and Figure 15 shows the most common individual BCTs. The 'Feedback and monitoring' group was the most common, with 92.3% of apps containing at least one BCT of this group, most commonly 'Feedback on behaviour' and 'Feedback on outcome(s) of behaviour' BCTs. 'Goals and planning' ('Goal setting' and 'Action Planning' BCTs) were also well represented at 84.6%. More than half of the apps included BCTs from the 'Comparison of behaviour' group (66.2%) of which the most common was 'Demonstration of the behaviour'. 'Social support' (64.6%), 'Shaping knowledge' groups (60%), and 'Associations' (46.2%) were common but only one BCT from each of these groups were present. The 'Reward and threat' group (53.8%) was common with two BCTs only ('Social reward' and 'Non-specific incentive'). Other BCT groups were rare: less than 15% of apps contained BCTs from the 'Comparison of outcomes' group; 'Natural consequences' and 'Antecedents' represented 10.8% and 6.2% of the total BCTs, respectively. The remaining BCT groups were non-existent in the PA apps.

Figure 14: Percentage of apps that included BCTs, presented by groups

Frequency of BCTs incorporated by PA apps, presented by BCT groups

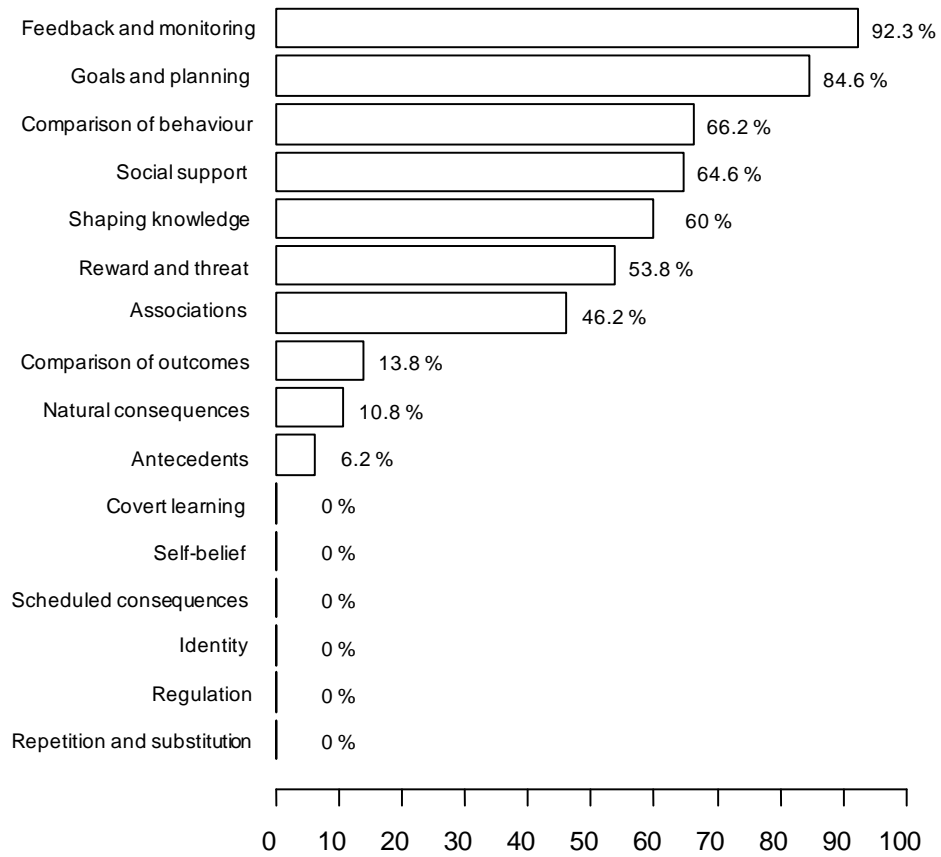
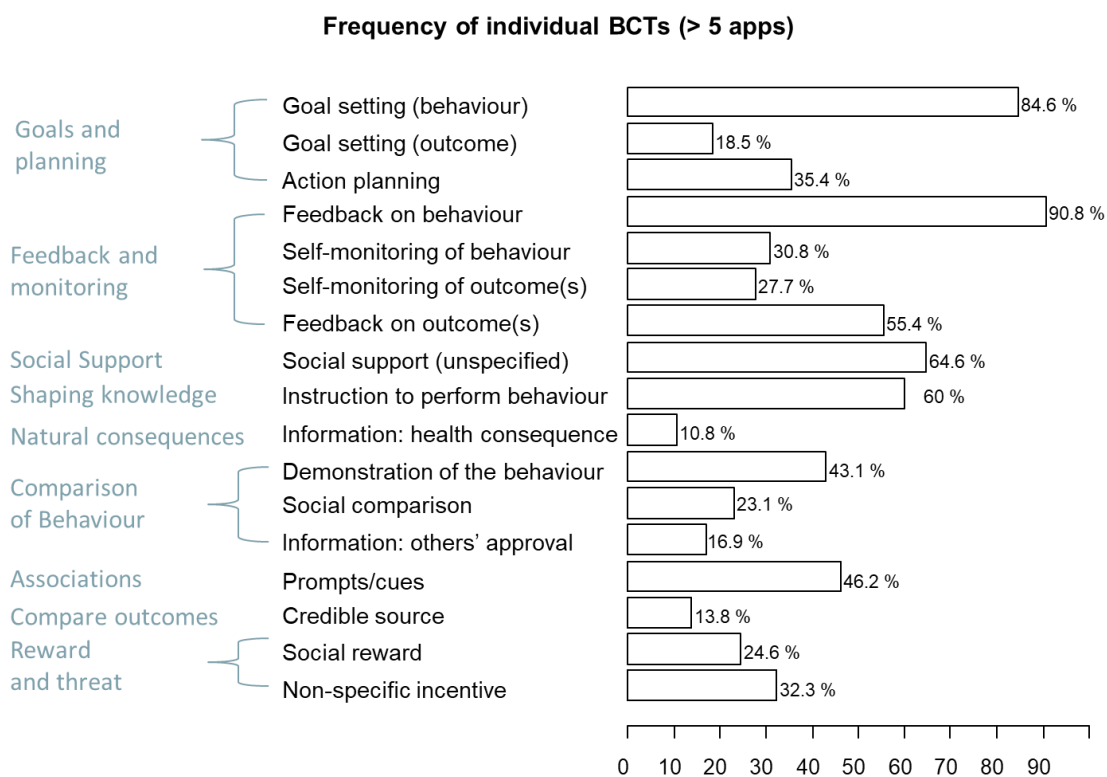


Figure 15: Frequency of individual BCTs within the groups BCTs (BCTs that occurred in at least five apps are shown)



### 2.4.2.5 App features

The Table 14 below displays the descriptive statistics for the app features extracted

Table 14: Descriptive statistics for the app features

		<b>Free (N=32)</b>	<b>Paid (N=32)</b>	<b>Total (N=64)</b>	<b>P value</b>
Number of Features	Mean ± SD	6.13 ± 2.73	6.44 ± 2.66	6.28 ± 2.68	0.645
	Median	6.50	7.00	7.00	
	25, 75%tile	4.00, 8.00	5.00, 8.00	4.00, 8.00	
	Min, Max	1.0, 12.0	1.0, 11.0	1.0, 12.0	



#### 2.4.2.6 App features and mapping of the features and the BCTs

There were 10 major features identified, and each of these features were divided into subgroups. Subsequently, the sub-features were mapped onto the BCTs (see Table 15 for the features and BCTs mapped).

The first major feature in the PA apps sample reviewed was related to *provision of information and guidance on PA* that were presented in written form, using images, videos or audio instructions. The most common BCTs expressed by this feature were 4.1. Instructions of how to perform the behaviour and 6.1. Demonstration of the behaviour.

The second and the third major features were related to feedback and self-monitoring. The major feature *Automatic tracking* relates to various methods the app collects information automatically, for example, using pedometers, accelerometers and timers. The difference between *Automatic tracking* feature and *Logs* is that the logs require the user to input the data by themselves, for example users input their weight regularly in order to monitor it over time. The most common BCTs expressed by the *Automatic tracking* feature were 2.2. Feedback on behaviour. BCTs related to self-monitoring (2.3. Self-monitoring of behaviour and 2.4. Self-monitoring of outcome(s) of behaviour) were expressed by the *Logs* feature.

The fourth major feature identified were *Reports*. These were based on either automatic tracking, for example written reports on completed PA sessions over the week/month etc., or graphs on number of steps. Example of *Reports* based on logs include summaries of user-inputted gym logs or visualisations of weight change over time. As the function of the reports is to give feedback to users on their performance of PA, the BCTs expressed by the *Reports* features were 2.2. Feedback on behaviour and 2.7. Feedback on outcome(s) of behaviour.

The fifth major feature found was *Gamification*. The examples of gamification included providing medals, badges and points for completed PA sessions; unlocking new content of features that are dependent on reaching a certain PA goal; taking part in various competitions where users can compare their performance. The most common BCTs that were expressed by the *Gamification* feature were 10.6. Non-specific incentive or 10.3. Non-specific reward. These were present when medals, badges, unlocking new content etc. was provided. The BCT 6.2. Social Comparison was observed in a feature where users had an option to enter PA competitions. The BCTs 1.1. Goal setting (behaviour), 1.3. Goal setting (outcome), and 1.4. Action planning were present for individual challenge feature.

The sixth feature, *Notifications* were divided into sub-features that reminded the users about their PA session. The BCT 7.1. Prompts/cues was expressed by the PA reminders sub-feature. Other notifications were also found, such as alerts when a new post was published on the forum. No BCTs were assigned for other notifications as they did not directly relate to PA behaviour.

The seventh major feature, *In-app community*, related to connecting with other users of the app. The sub-features included forums where users could share experience and advice on PA. This feature also included facilitated search of real-life PA events, and access to a database of user-created PA sessions, for example running routes where other users could compete for time.

The most frequent BCTs that were expressed by the *In-app community feature* were 3.1. Social support (unspecified), 3.1. Social support (practical), 6.3. Information about others' approval, 6.2 Social Comparison.

The eighth major feature, *sharing on social media*, related to the ability to connect with other users through posting of information on completed PA sessions on various social media, most notably Facebook and Twitter. The most frequent BCT s expressed by this feature was provision of 3.1 Social Support (unspecified).

The ninth major feature was *Integration with other apps or devices*. Some of the apps in the sample had an option to connect to other apps to gain health-related data or to access music library in order to facilitate a running session. Some apps had an option to connect to external sensors such as heart rate monitors No BCTs were mapped for this feature.

The last major feature, was *tailoring of PA*. This feature related to the potential for personalising the settings of the app to the user preferences or needs. The tailoring related to the ability to choose the type, duration, pace, and timing of PA sessions. No BCTs were mapped for this feature.

Table 15: Major feature group and subgroups mapped with BCTs

	<b>Major feature</b>	<b>Features subgroups</b>	<b>BCTs “expressed” by the features in the study</b>
1.	PA EDUCATION IN A VARIETY OF FORMATS  <i>Definition: information, instruction, guidance on PA presented as written material, image, audio or video instructions.</i>	Written material	4.1. Instruction on how to perform the behaviour 5.1. Information about health consequences 12.1. Restructuring the physical environment; 12.2. Restructuring the social environment
		Images	6.1. Demonstration of the behaviour

	Major feature	Features subgroups	BCTs “expressed” by the features in the study
		Videos instructions (including animation)	6.1. Demonstration of the behaviour
		Audio instruction	4.1. Instruction on how to perform the behaviour; 12.1. Restructuring the physical environment; 12.2. Restructuring the social environment
2.	AUTOMATIC TRACKING (passive data collection)  <i>Definition: any data automatically recorded by the app.</i>	Pedometer (step count)	2.2. Feedback on behaviour; 1.1. Goal setting (behaviour)
		Accelerometer or GPS (distance, duration, average speed, distance alert, e.g., when 1km was reached, mapping the route)	2.2. Feedback on behaviour
		Timer (measuring the time spent active or exercise set completion)	2.2. Feedback on behaviour; 1.1 Goal setting (behaviour)
		PA session completion tracker (based on timer or user feedback, e.g., tapping when certain set of exercise finished)	2.2 Feedback on behaviour; 1.1 Goal setting (behaviour)
		Estimation of calories burnt; BMI	2.7. Feedback on outcome(s) of behaviour
3.	LOGS (active user-inputted self-monitoring)	PA performed logged by the user	2.3. Self-monitoring of behaviour; 1.1 Goal setting (behaviour)
		Calories; weight; photos of oneself taken over time	2.4. Self-monitoring of outcome(s) of behaviour

	Major feature	Features subgroups	BCTs “expressed” by the features in the study
	<i>Definition: any data inputted by the user.</i>		
4.	REPORTS	Report based on automatic tracking (written reports, e.g., summaries of PA session completed in a day/week/month, etc.; visualisations, e.g., graphs of number of step count or distance run over week/month, etc.)	2.2. Feedback on behaviour; 2.7 Feedback on outcome(s) of behaviour
		Reports based on logs, i.e., user-inputted data (written reports, e.g. summaries of user-inputted gym logs; visualisations, e.g., graphs of weight change)	2.2. Feedback on behaviour; 2.7 Feedback on outcome(s) of behaviour.
5.	GAMIFICATION*	Achievements, medals, badges, points, scores, stars	10.6. Non-specific incentive; 10.3. Non-specific reward
		Unlocking new content or features dependant on reaching a certain PA goal e.g., unlocking new exercise; unlocking different skins, i.e., graphical appearance	10.6. Non-specific incentive; 10.3. Non-specific reward; 1.1. Goal setting (behaviour); 1.3. Goal setting (outcome); 1.4. Action planning
		Unlocking new exercise dependant on reaching a PA goal when the exercise would have to be paid for otherwise	10.1. Material incentive (behaviour); 10.2. Material reward (behaviour)
		Competitions e.g., Levels of achievement or rank; Leaderboards;	6.2. Social comparison

	Major feature	Features subgroups	BCTs “expressed” by the features in the study
		Challenges (individual)	1.1. Goal setting (behaviour); 1.3. Goal setting (outcome); 1.4. Action planning
		“Story running” (audio story where a runner is the main character)	-
6.	NOTIFICATIONS	PA session reminders	7.1. Prompts/cues
		Other notifications	-
7.	IN-APP COMMUNITY	Forum (also called feed or club) used for sharing experiences and advice on PA; cheering others while they doing the PA sessions.	3.1. Social support (unspecified); 6.3. Information about others’ approval
		PA real world event finder that facilitates easy sign up	3.1. Social support (practical)
		User-created PA sessions (e.g., running routes where other users can compete for time)	6.2. Social Comparison
8.	SHARING ON SOCIAL MEDIA	Posting information on completed PA session on social media	3.1. Social support (unspecified)
9.	INTEGRATION WITH OTHER APPS OR DEVICES	Apple Health	-
		Music library	-
		Playlist on Spotify	-
		Sensors (e.g., heart rate monitor)	-
10.	TAILORING of PA	Setting type, duration, pace, or timing of PA sessions	

*Note:* \*Gamification was defined as “the use of game design elements in non-game contexts”[186]. The six core components of gamification were taken from the review of gamification elements used in health apps [160]. These were:

- (1) leaderboards,
- (2) levels of achievement or rank,
- (3) digital rewards (e.g., points, medals, badges),
- (4) real world prizes,
- (5) competitions/challenges,
- (6) social or peer pressure.

#### 2.4.2.7 Quality of app development and evaluation process

Only 1 app had a non-commercial affiliation, *One You Couch to 5K*, which was developed by PHE (see Table 16). None of the apps reported user involvement during development. Twelve out of 65 apps (4 free and 8 paid) consulted with experts to design the content of the app. Nine out of 23 free apps (28.1%) had a study associated with the apps published in a peer-reviewed journals. In comparison, for only 3 paid apps (9.1%) there was a peer-reviewed study found.

Table 16: Descriptive data for the quality of app development and evaluation process: organisational affiliation, expert and user involvement, and evidence of evaluation in peer-reviewed journals

		Free (N=32)	Paid (N=33)	Total (N=65)	p- value
Affiliation	Commercial	31 (96.9%)	33 (100.0%)	64 (98.5%)	0.492
	Government institution	1 (3.1%)	0 (0.0%)	1 (1.5%)	
Any Expert	No	28 (87.5%)	25 (75.8%)	53 (81.5%)	0.339
	Yes	4 (12.5%)	8 (24.2%)	12 (18.5%)	

Any User Involvement	No	32 (100.0%)	33 (100.0%)	65 (100.0%)	
Any Peer Journal	No	23 (71.9%)	30 (90.9%)	53 (81.5%)	0.061
	Yes	9 (28.1%)	3 (9.1%)	12 (18.5%)	

## 2.4.2.8 User experience

### 2.4.2.8.1 User ratings

#### The highest-rated apps

The highest-rated free apps on iTunes was *Sworakit* developed by *Nexercise* (star average of 4.74), while the highest-rated paid app in the same store was *Couch to 5k Runner* developed by *Fitness22 Ltd.* (star average of 4.86). In GP the highest-rated free app was *Interval Timer* developed by *dreamspark* (star average of 4.73), and the highest-rated paid app was *Fitness Trainer FULL version* by *Your personal coach* (star average of 4.70). For one app, *Yoga Break* by *Fitness Guide Inc.*, the data on user ratings were not available due to unreported number of ratings.

The median user rating in iTunes was 4.4 and 4.5 in GP and did not differ between free and paid apps in either stores.

Descriptive data for user ratings are presented in Table 17 for iTunes and 18 for GP. The median user rating was 4.4 stars and did not differ between free and paid apps in iTunes.



Table 17: descriptive statistics for user ratings in iTunes

		<b>Free - iTunes (N=21)</b>	<b>Paid - iTunes (N=24)</b>	<b>Total - iTunes (N=45)</b>	<b>P- value</b>
Avg. user rating (1-5 stars)	Mean ± SD	4.1 ± 0.8	4.3 ± 0.6	4.2 ± 0.7	0.251
	Median	4.4	4.6	4.4	
	25, 75%tile	4.0, 4.6	4.0, 4.8	4.0, 4.6	
	Min, Max	2, 5	3, 5	2, 5	

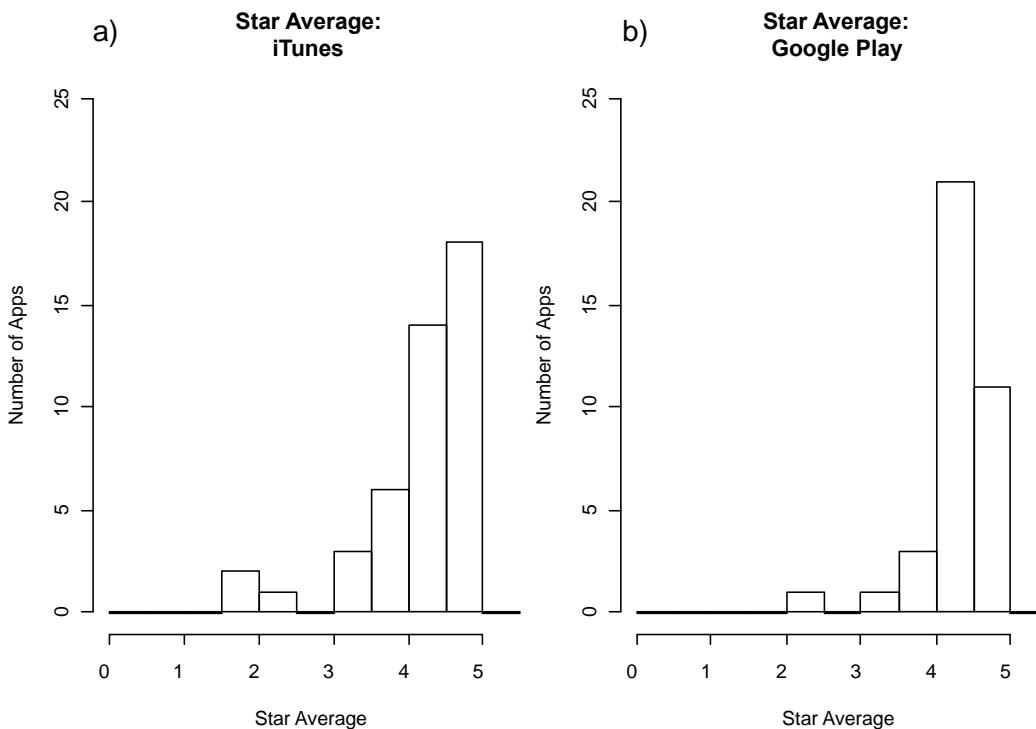
The user rating in GP was similar to those in iTunes with 4.5 median star average and did not differ between free and paid apps.

Table 18: descriptive statistics for user ratings in GP

		<b>Free - GP (N=21)</b>	<b>Paid - GP (N=16)</b>	<b>Total - GP (N=37)</b>	<b>P- value</b>
Avg. user rating (1-5 stars)	Mean $\pm$ SD	4.4 $\pm$ 0.5	4.4 $\pm$ 0.3	4.4 $\pm$ 0.4	0.784
	Median	4.5	4.5	4.5	
	25, 75%tile	4.4, 4.6	4.4, 4.6	4.4, 4.6	
	Min, Max	2, 5	4, 5	2, 5	

In both stores, the 25th percentile was around 4 stars (4.0 in iTunes and 4.4 in GP) suggesting that the user ratings were quite high and only 25% of ratings were below 4 stars. The histograms of star ratings in both store (Figure 16) showed the skewness of the star average distribution.

Figure 16: Distribution of user ratings in iTunes and GP



### 2.4.2.8.2 SUS

The results of the SUS assessment are summarised in Table 19.

Table 19: Descriptive data for the SUS assessment

		<b>Free (N=32)</b>	<b>Paid (N=33)</b>	<b>Total (N=65)</b>	<b>p- value</b>
SUS Total - PB	Mean ± SD	75.5 ± 17.8	80.5 ± 15.4	78.1 ± 16.7	0.232
	Median	77.5	82.5	80	
	25, 75%tile	62.5, 90.0	75.0, 87.5	72.5, 90.0	
	Min, Max	30, 100	25, 100	25, 100	
SUS Total - GA	Mean ± SD	87.0 ± 14.2	90.4 ± 17.1	88.7 ± 15.7	0.383
	Median	92.5	97.5	97.5	
	25, 75%tile	77.5, 97.5	90.0, 100.0	82.5, 100.0	
	Min, Max	45, 100	28, 100	28, 100	
SUS Total Score Difference	Mean ± SD	-11.4 ± 19.8	-9.8 ± 22.2	-10.6 ± 20.9	0.767
	Median	-6.3	-10	-10	
	25, 75%tile	-22.5, 2.5	-20.0, -2.5	-20.0, 0.0	
	Min, Max	-60, 18	-73, 60	-73, 60	
SUS Average Score	Mean ± SD	81.3 ± 12.6	85.5 ± 11.9	83.4 ± 12.4	0.172
	Median	85	87.5	86.3	
	25, 75%tile	71.9, 91.3	80.0, 93.8	75.0, 92.5	
	Min, Max	53, 100	58, 100	53, 100	

Note: SUS, System Usability Scale

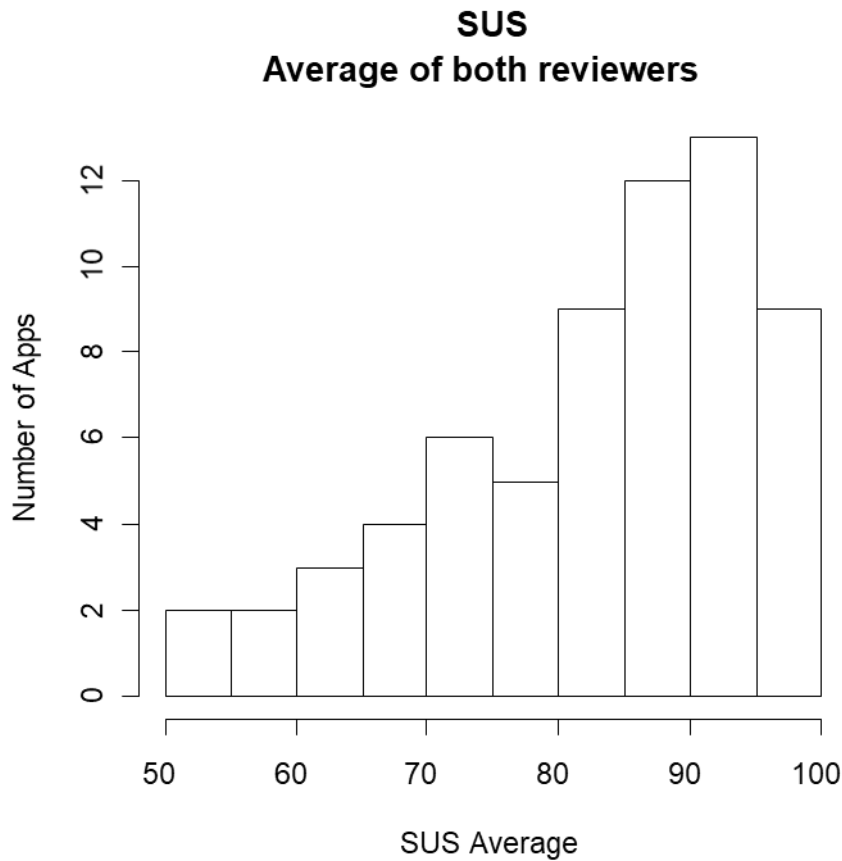
### **Total score difference between the reviewers**

The total SUS score for the first reviewer (PB) had a median of 80.0. Fifty percent of the SUS assessment for the sample had a score between 72.5 and 90. For the second reviewer (GA), the median score was 97.5 and typical scores fell between 82.5 and 100. The median SUS score difference between the reviewers was 10.0. In 50% of the cases, PB's SUS ratings were between 20.0 to 0 scores lower than those of GA. 25<sup>th</sup> percentile difference was 20 and the for 75<sup>th</sup> percentile there was no difference in ratings. The difference in SUS score between the two reviewers suggests that the first reviewer (PB) gave consistently lower scores than the second reviewer (GA), across all apps. The means were statistically significantly different with mean difference of -13, 95% CI 18.9 to -7.1,  $p < 0.0001$ .

### **SUS score of the sample**

The average SUS score for the apps was similar for both free and paid apps, with median of 86.3. Using descriptors (explained in Figure 5) [165], the score can be described as excellent. 50% of the total average SUS score fell between 75.0 and 92.5 and 25% had a score higher than 92.5, suggesting that more than 75% of the app sample assessed could be described as having good to excellent usability. Figure 17 presents the total SUS score averaged for the app sample.

Figure 17: Total SUS score



## 2.5 Discussion

### 2.5.1 Principal findings

This study described the most popular PA apps on the market, focusing on the quality determinants of safety (data privacy and security), effectiveness (BCTs and development and evaluation quality), and provision of the most positive experience possible (user ratings and usability). Overall, the findings suggest that most of the apps in this sample were of reasonable quality in terms of the user experience but there were substantial shortcomings in the areas of safety and effectiveness. The assessment of data privacy and security showed that the privacy policy was not available for 29.2% of the apps. Most apps collected PII, shared users' data with 3<sup>rd</sup> party, and more than half of the apps did

not specify how they ensure data security. Every app contained at least 1 BCT, with an average of 7. The maximum number of BCTs was 13 and the most common BCTs related to provision of feedback on behaviour. All but 1 app had commercial affiliation, 12 consulted an expert and none reported involving users in the app development. Only 12 of 65 apps had a peer-reviewed study connected to the app but only 1 app was assessed for efficacy in a trial [187]. User ratings were high with only a quarter of the ratings falling below 4 stars. Similarly, the usability scores were “good” to “excellent”. There was no statistically significant difference between free and paid apps on the characteristics or quality indicators.

### **2.5.2 Safety of apps**

The assessment of privacy policies showed that privacy and security of users’ data could be substantially improved. These results are consistent with previous studies assessing data safety. Huckvale, Prieto [60] who assessed the apps from the NHS Apps Library, found that 20% of who assessed the apps from the NHS Apps Library, found that 20% of apps did not have a privacy policy, and most of the apps breached users’ data privacy and security.

Powell, Torous [153], who assessed inter-rater reliability of various app assessment measures, found only moderate agreement on the measure of data privacy and security. This result might be somewhat surprising considering that the assessment is factual rather than subjective in nature and should therefore reduce the likelihood of different interpretations. However, the findings of this study support these results as most of the policies used vague legal jargon that could be difficult to comprehend for a potential user.

Collecting and analysing consumer data by app developers can have advantages for the users, such as personalization and improvement of the products [35]. However, the information about these practices ought to be transparent and understandable [36] to

enable the potential user to make an informed decision to download the app. Regulatory oversight concerning data protection is challenging because of the large scale of the app market. In consequence, ensuring the privacy and security of data is left in the hands of app developers [126]. It will be of interest to investigate the effect of the General Data Protection Regulation on availability and accessibility of the privacy policy, and on the data safety and data sharing practices of app developers.

### **2.5.3 Likelihood of effectiveness**

The apps in the review contained, on average, 7 BCTs. The results of this study are similar to those found in previous reviews of PA apps: Middelweerd, Mollee [145] found, on average, 5 BCTs were used in each app; Conroy, Yang [167] reported between 1-13 BCTs with a mean of 4.2; study using the same BCT Taxonomy as the one in this study found , on average, 6.6 BCTs [139].

The most common BCTs present in apps in this study were goal setting and action planning, feedback and monitoring and these findings are in-line with previous reviews and content analyses of apps [137, 145, 167, 176]. These BCTs can be defined as self-regulation strategies. Indeed, increasing self-regulatory capacity using tools such as mobile devices is convenient as the available features enable setting goals, self-monitoring, and provision of feedback. In the sample assessed, information on the activity performed could be logged manually but, most often, the app would automatically record the PA using GPS (running, walking, cycling) or timers (completion of a workout). These data were used to produce feedback, either on behaviour or on outcomes of behaviour, in a form of engaging visualisations.

Self-regulation strategies have been shown to be effective in increasing PA behaviour [140, 141, 188] and may show larger effect if presented together. For example, Michie, Abraham [140] showed that self-monitoring with at least 1 other self-regulatory BCT (e.g.,

goal setting, action planning, and feedback) was more effective at increasing PA than using only 1 on these BCTs in isolation. However, self-regulation is a limited resource which requires deliberate effortful decision-making process [189]. A review of theories concerned with maintenance of behaviour identified five theoretical themes relevant to behaviour change maintenance. Although, self-regulation techniques are crucial for the maintenance of behaviour change, four other factors play an important role in successful behaviour change: 1) motives for maintenance, such as self-determination. Initial behaviour change is often motivated by extrinsic motivation, maintenance of behaviour requires intrinsic motivation (i.e., the behaviour is congruent with one's values and identity); 2) Habit formation which enables to develop automaticity for the new behaviour so that relapse can be prevented during periods of low resources when self-regulation is prone to failure; 3) Resources, both physical and psychological, to enable sustained effort; 4) Environment and social influences which provides incentives and disincentives to behaviour change. Hence, app features beyond self-regulation techniques should be considered in PA behaviour change interventions delivered using apps.

Beyond the self-regulation techniques, the evidence of what BCTs are most likely to increase the likelihood of behaviour change is unknown. A recent meta-analysis of online interventions suggests that certain BCTs are more efficacious when present together [190]. These results were based on studies not specific to PA apps and it is likely that different BCTs may be more suitable for different modes of delivery (face to face, web-based, app), e.g. social support might produce better results when delivered face to face rather than via an app. On the contrary, automatic monitoring and feedback on PA in apps can facilitate self-regulation and may be considered as a more efficient method than self-recording using diaries.



The findings show that BCTs from 9 out of 16 BCT groups were rare or non-existent in the apps assessed. Twenty-five of 93 possible BCT were used which constitutes 27% of the current BCT Taxonomy. The evidence of what BCTs are most likely to increase the likelihood of behaviour change is unknown. It is possible that certain BCTs are more efficacious when present together producing a synergistic effect [190].

Similarly, the effect of the number of BCTs on efficacy of the interventions remains inconclusive. There is some evidence that higher number of BCTs produces larger effect sizes in web-based interventions [191], others studies show no effect [140].

The use of a variety of BCTs groups as well as the techniques within the BCT group would theoretically increase effectiveness by addressing various barriers to PA. For example, within the 'Goals and planning' BCT group, only 3 out of 9 BCTs were utilized. Implementing features that utilize other BCTs that enable goal-setting and planning (e.g., problem-solving technique, asking user to commit to their goal and providing opportunity for the user to review their goal) might increase the likelihood of effectiveness of the app.

The use of evidence and theoretical frameworks is vital in developing behaviour change interventions [134]. COM-B model (Chapter 3) of behaviour change [114] enables to systematically identify the barriers and facilitators of the behaviour targeted, and to select intervention components that will address these barriers in order to increase the likelihood of behaviour change.

The findings suggest that the quality of the app development and evaluation process could be improved. There was no evidence of user involvement and most apps were commercially developed with rare involvement of health or behaviour change experts. Similar results were found in previous reviews [137, 169, 192] and there is evidence to

suggest that expert involvement predict the number of app download [193]. Indeed, the user-centred design framework stresses the importance of understanding the contextual experiences of potential users, as well as inclusion of multidisciplinary skills and perspectives when developing products and services. Our results also support previous research showing the lack of evidence for scientific evaluation of the apps on the market [194, 195]. We found only 12 studies in peer-reviewed journals that were associated with the apps. However, only 1 app was used in a pragmatic RCT [187] and the study was not conducted by the app developer.

#### **2.5.4 Positive experience**

The usability of the apps reviewed was high. Likewise, user ratings of the PA apps were high with only a quarter of the ratings receiving less than 4 stars. Similarly, Mendiola, Kalnicki [147] found that usability was related to user ratings in a general sample of health apps. The competition for customers in the app stores is high with 90% of apps in the app stores not attracting enough attention to feature in the ranking of the app stores and are consequently not visible for the user, called “App Zombies” [196]. High quality graphic design, visual appeal and ease of use are more likely to attract potential customers to download and engage with the app. However, it is unknown whether these variables relate to effectiveness of the apps. There is evidence to suggest that web-based interventions with higher usability tend to be more effective [190]. However, continued engagement with an app may suggest engagement with the intervention or unhealthy dependence [197].

#### **2.5.5 Strengths**

The strengths of this study include a systematic approach to sample identification and assessment. First, the sample of apps was identified by screening 400 apps in two major app distribution platforms including both free and paid apps. Second, the sample was identified and assessed by two independent reviewers. Third, the assessment tools

covered various aspects of quality: inclusion of the BCTs, as well as user experience using subjective (user ratings) and objective (usability) measures. Last, this was the first attempt to describe PA app features and map these with BCTs.

### **2.5.6 Limitations**

First, it is unknown what variables are included in the ranking algorithm of the top apps from which the sample was selected. It is likely that usage data and user ratings comprise the ranking [198] but other unknown variables may also be included. Second, the possibility that user ratings were influenced by fake reviews cannot be excluded. [148, 149]. However, there is a reliance on genuine users of the app to mark it down if the app does not live to their expectations and this review included popular apps with high number of ratings (2.8 million). Third, data privacy and security assessment was limited to the analysis of the policy. There is evidence of inconsistency between the policy statement and the actual practices of app developers [60]. Fourth, the intensity of implementation of each BCT was not assessed in the present study. Fifth, the quality of app development process was based on the information provided in the app stores, the app website and in the app itself), hence it is possible that some data were missed if they were not available online. Sixth, as this is the first attempt to map the features with BCTs, the list is not an exhaustive but limited to the sample assessed. There were BCTs that were not matched with the features, and vice-versa. Last, as none of the app developers claimed to have theoretical basis underpinning the development of the app features, the extracted BCTs were based on the experienced content of each app as assessed by two reviewers.

### **2.5.7 Implications**

More research is needed to understand the use of PA apps in order to design effective digital tools. There is little knowledge concerning how users adopt these apps into daily routines, what are the facilitators and barriers to increasing PA using apps. Secondly,

the optimal number of BCTs in PA apps is unknown, neither is the combination of BCTs that is likely to be effective in increasing PA facilitated by app.

Although popularity of the apps is high, healthcare professionals and potential users need to be aware of the limitation in the safety of personal data as well as the limitation in the quality of the apps to change behaviour. Currently, it is not possible to recommend apps that are most effective but attempts to create database of high quality apps are in progress.

Based on the results of the data privacy and security assessment, recommendations for best practice include: 1) app providers should include a clear statement, written in plain English, which summarises the main points of the privacy policy, 2) this statement should direct users who wish to know more to the detailed policy. In future, assessments of data privacy and security could use the tool used in this study.

A multidisciplinary collaboration between app developers and behaviour change experts could increase the potential of PA apps to change behaviour. Behaviour change experts can provide input into the content of apps that will increase the likelihood of behaviour change, while app developers can provide creative input into how to make these apps more engaging, and hence used. This, in turn, may increase the potential of mobile tools to improve population health. In order to facilitate communication between collaborators from different fields of expertise, the list of features designed in this study can provide a tool to help design the content of the digital interventions as it mapped the common features with the theoretically derived behaviour change techniques.

### **2.5.8 Conclusions**

This study examined the quality of the most popular PA apps currently available on the market. Although usability and user ratings of apps were high, there was a concerning

lack of safety controls for users' personal data for the majority of the apps, the apps included limited number of BCTs that mostly related to feedback on behaviour, and the quality of the content and development processes were suboptimal. The technological development and the potential for profit far outpaced the research on the ability of these apps to support PA behaviour change. With 165,000 on the market, this represents a loss of opportunity for health promotion on a large scale.

### **2.5.9 Citation for the published peer-reviewed article for this study**

Bondaronek, P., Alkhalidi, G., Slee, A., Hamilton, F. L., & Murray, E. (2018). Quality of publicly available physical activity apps: review and content analysis. *JMIR mHealth and uHealth*, 6(3), e53. <https://doi.org/10.2196/mhealth.9069>

See Appendix R for the published peer-reviewed journal article for Study 1.

## CHAPTER 3. The relationship between popularity and app quality: regression models

### 3.1 Chapter overview

**Background:** PA apps are the largest category of health apps available in the major app stores, suggesting substantial public interest in increasing PA. However, research into their effectiveness is scarce, and studies assessing the association between the popularity of these apps and their quality are limited. The user-led feedback in the form of star ratings assigned to the apps can potentially provide some insights into the user experience of these apps.

**Aim and objectives:** The aim of this study was to explore which factors predict higher user rating in the app stores. The primary objective of interest was to assess the association between the number of BCTs and user ratings. The secondary objective was to assess the quality indicators and app characteristics associated with BCTs and user ratings.

**Methods:** 400 top-ranked free and paid apps from iTunes and GP stores were screened, and were included if: (i) the primary behaviour targeted was PA; (ii) they had stand-alone functionality; and (iii) they targeted adults (see Chapter 2 for the details of the sample). The outcome variable of user rating was dichotomised into high (4, 5 stars) or low (1, 2, 3 stars) ratings. Logistic regression models were performed to determine the association between the presence of BCTs and user ratings. Variables determined a priori as potentially affecting the relationship between BCTs and user ratings were app store (iTunes or GP), number of features, whether the app was free or required payment, app size and usability.

**Results:** There was no relationship overall between star ratings and the BCTs. Sub-group analysis looking at the relationship between star ratings and number of BCTs by store found that in iTunes there was an association with each additional BCT corresponding to 15% increase in the chance of a higher star rating (OR: 1.15, 95% CI: 1.06, 1.25,  $p= 0.001$ ). App store was strongly associated with star ratings, with iTunes users less likely to give 4 or 5 stars compared to GP users (OR: 0.74 95% CI: 0.73-0.76,  $p < 0.001$ ). Predictors of the number of BCTs that were found significant were app size, the number of app features, the presence of peer-reviewed studies associated with the app, and the ranking in the apps stores. Expert involvement predicted higher user ratings in iTunes, but not in GP. There was no other associations between the app quality indicators or characteristics and user ratings.

**Conclusions:** This is the first study to utilise data extracted from the content of the apps and app stores to predict higher user ratings in highly-ranked PA apps. The sample investigated in this study were highly rated, highly-ranked apps from the major app stores, hence most visible for the potential user. However, this research suggests that popularity does not necessarily imply high quality and likelihood of effectiveness. Hence, public health impact is unlikely to be achieved by allowing market forces to “prescribe” what is used by the public.

## 3.2 Background

As discussed in Chapter 1, engagement with apps is low, with up to 90% of downloaded apps only used once [147] and long-term use, likely to be needed for effective behaviour change, is relatively rare: 68% of users open less than 5 apps at least once a week [79]. Although prolonged engagement with digital interventions may indicate dependence [197], there is a consensus that engagement is needed for intervention effectiveness [200]. Hence, it is important to shed light on what app content and characteristics might be popular and valued by the users.

App stores provide user-led app feedback systems that can potentially provide a rich source of information about the use of the apps [146]. The user rating measure in the form of star ratings is an imperfect assessment of actual user experience, however, in the absence of the evidence of what users value in health apps, the star ratings may provide some insight into user experience of health apps, while being convenient to obtain.

Hence, whilst Study 1 investigated the quality of the sample of PA apps from the perspective of a researcher, this study focused on the voice of the users as expressed by the app user ratings to understand what users value in popular apps. Specifically, this study will utilise the data extracted in Study 1 to shed light on the app content and characteristics that might be valued by users. First, I describe, evaluate, and classify the studies that investigated the relationships between the variables derived from the app stores focusing on user experience and the quality of app content. I then move on to discuss the methodology used to assess the association between quality indicators and other app characteristics in the sample of the apps described in Study 1.



### ***3.2.1 Research exploring the factors that influence user experience based on the data derived from the app stores***

Studies that explored the association between the user experience and data derived from the app stores focused on user ratings, ranking in the apps store, and number of downloads as proxies for user experience measures.

Harman, Jia [146] and Pustozarov, von Jan [201] used data mining, a method for identifying patterns in large data sets. Harman, Jia [146], as one of the first to explore the potential of using data available in the app store, extracted information from 32,108 apps and assessed the association between app features, number of downloads, price and user rating. They found a correlation between user ratings and the number of downloads. They used data derived from the BlackBerry app store in 2011 which is now obsolete, having only 0.05% of the overall phone market share [202]. In addition, the features of apps assessed were broad and focused on all apps, rather than health apps specifically.

More recently, Pustozarov, von Jan [201] assessed 46,430 apps from the medical and health and fitness category of the GP store. They found the frequency of assigning a 4 star rating was positively associated with the number of downloads. There was also a correlation between the last update and user ratings, with ratings steadily decreasing with longer time since the last update. They also found that the most downloaded apps were less likely to receive five stars than the ones with low download rates. This might suggest that apps that are highly downloaded are rated by users rather than companies that specialise in fake user ratings, a concern that has been proposed as a limitation of drawing conclusions based on user ratings. However, the authors also found correlations that they described as “unexpected”, such as the association between user ratings and the app vocabulary and descriptions in the app store. Specific vocabulary and

descriptions found to be correlated with the ratings were not provided and the authors did not respond to a request for further information. However, it is possible that when such a high number of apps are assessed associations between different characteristics and user rating could occur by chance. This example illustrates the importance of using an analysis plan based on theory to assess the relationships between the variables available in the app stores.

Mendiola, Kalnicki [147] aimed to determine which app features and characteristics predicted higher user rating in the app stores. Their app sample included 234 apps derived from the health apps category, which includes various well-being and disease management apps. They found that no single variable significantly impacted on star rating. In their multiple regression analyses the best model showed that 9.3% of ratings could be explained by 5 features and app characteristics. Mendiola, Kalnicki [147] showed that providing features that enable plan of action, tracking and the ability to send data to healthcare providers had a positive impact on user rating. Apps that were more usable and required payment were rated higher than those that were less usable and free. Arnhold, Quade [203] assessed 656 apps targeting the management of diabetes. Similar to Mendiola, Kalnicki [147] they assessed a subset of 66 apps for their usability and found that the number of different app functions (from the list of 7) was negatively correlated with usability, suggesting that more complex apps might be less intuitive and more difficult to use.

Both studies focused on a limited number of features. Arnhold, Quade [203] extracted 7 broad app functions. For example, the “information function” would most likely include any type of information provision. However, the medium of information delivery: written text, videos, pictures, will also influence the user experience of the app [152]. Mendiola, Kalnicki [147] included 12 features that were “aligned with” self-determination theory

[204], however, the process of theory application to identify those features was not reported. It is crucial for digital health interventions to be grounded in behaviour change theory if they are to be effective [134, 205].

### ***3.2.2 Research exploring the relationship between the quality of app content based on the data derived from app stores***

Studies that explored the association between the app content and other variables derived from app stores focused on app features, various quality indicators such as scientific coverage, expert involvement, adherence to the theory of behaviour and behaviour change, and Mobile App Rating scale which is a researcher-led questionnaire assessing the quality of health apps [177].

Research assessing the relationship between the popularity of apps and their quality has found mixed results. Azar, Lesser [172] found that apps for weight management that were of higher quality, defined as inclusion of the constructs from four behaviour change theories, were not among the highest ranked apps in app stores. Similarly, apps that had higher download rates or higher rankings were associated with less adherence to guidelines in smoking cessation apps [206]. On the other hand, Pereira-Azevedo, Osório [193] reviewed 129 urology apps in the GP app store and found that higher download rates were associated with expert involvement in the development of the apps. These studies exploring the relationship between popularity and quality in apps need to be interpreted with caution. They targeted different health conditions and behaviours, and used different definitions of quality, such as consistency with behaviour change theory, expert involvement, and adherence to guidelines.

The Behaviour Change Technique (BCT) Taxonomy [116] is a systematic method used to specify the content of behaviour change interventions [113] and it has also been used to quantify the behaviour change content of interventions, including apps, [136, 137,

139]. The Taxonomy provides an efficient method of app content classification and enables comparison and aggregation of finding by providing a common language across behaviour change research literature, and was used to describe the change techniques in this study. A number of studies have used Behaviour Change Techniques (BCTs) and focused on assessing the association between the BCTs and data derived from the app stores.

Chen, Cade [137] reviewed 28 apps for weight loss and conducted regression analyses to determine the relationship between the popularity of the apps (operationalised as the ranking in the app stores) and the quality assessment criteria which included the presence of BCTs (using the CALOR-E taxonomy[168]), accountability, (such as developer's credentials and affiliation, provision of information sources and references, sponsorship disclosure), scientific coverage and content accuracy, as well as usability assessed using the System Usability Scale [162]. They found that app ranking was associated with total quality score, with scientific coverage and accuracy and with incorporation of BCTs. They also found positive correlation between the number of features in the apps and the ranking.

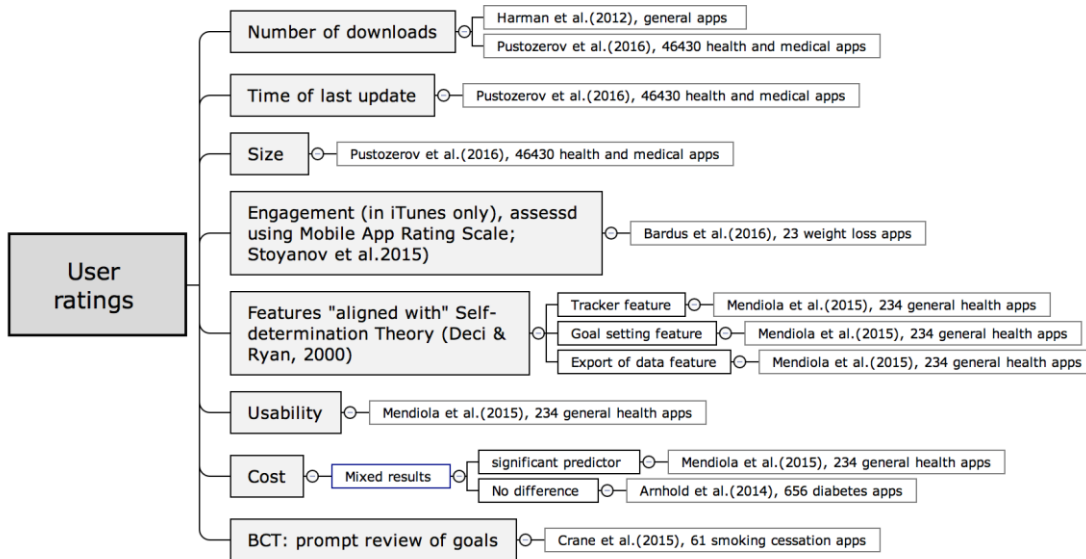
Bardus, van Beurden [136] reviewed 23 apps aimed at weight management. They used the MARS scale, which is a reviewer-led assessment of the quality of apps. The analysis showed a positive association between the number of BCTs and the *Engagement, Functionality, and Aesthetics subset*, as well as *the* overall MARS score. They also extracted 9 app features which showed a positive correlations with the number of BCTs.

Similarly, Crane [138] showed that only 1 BCT, “prompt review of goals”, was associated with user ratings in alcohol reduction apps. The authors concluded that there is little association between the mention of theory, BCT inclusion and the

popularity of the alcohol reduction apps. These findings suggest that further work is needed to explore whether popular apps, such as those aiming at increasing PA, are of high quality and likely efficacy. This has relevance for public health as the popularity and likely efficacy link creates vast potential for increasing PA on a population scale, hence preventing morbidity and premature mortality.

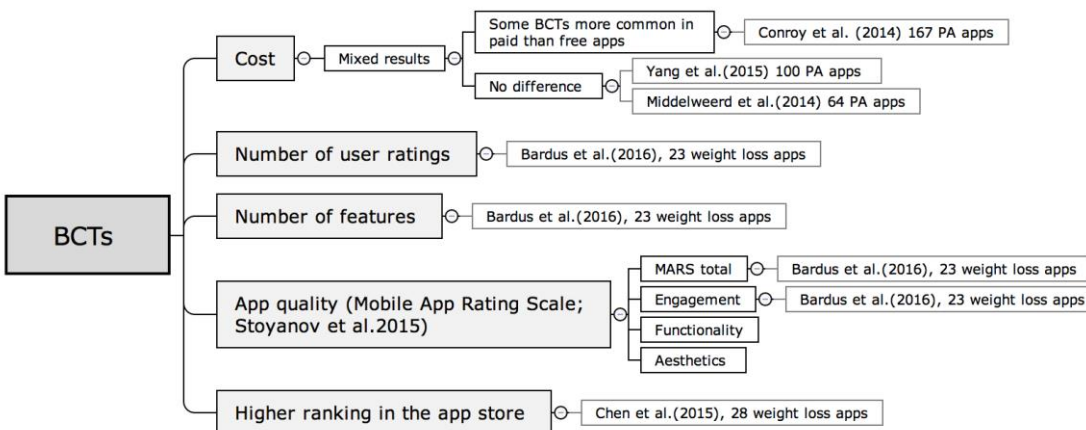
In sum, previous studies that investigated the indicators of user experience of apps as assessed using the “crowd-sourced” measures of user ratings and data derived from the app stores used various definitions of app quality, assessed a wide range of app characteristics, and targeted apps tackling different behaviours and health conditions. It is likely that users value different characteristics of apps depending on the health behaviour or health condition targeted and none of the studies focused on PA apps. Hence, this study used the quality indicators identified in Study 1 to assess the predictors of positive user experience, as assessed by user ratings in the highly ranked PA app sample. A visual representation of the significant associations found in studies that assessed user ratings and the inclusion of BCTs is presented in Figures 18 and 19.

Figure 18: The visual classification of the significant associations found in studies that assessed user ratings



Note: the associations indicated were positive between user ratings and number of downloads, size, app features, usability, BCT: prompt review of goal; negative between user ratings and time of last update. The mixed results are indicated on the diagram.

Figure 19: The visual classification of the significant associations found in studies that assessed the inclusion of BCTs



Note: the associations indicated were positive between the number of BCTs and the number of user ratings, number of features, app quality rating, and ranking. The mixed results are indicated on the diagram.

### **3.2.3 Aims and objectives**

The overall aim of this study was to explore what predicts higher user rating in the app stores. The specific objectives are listed below.

Objective 1:

- To assess the association between the number of BCTs and user ratings
- To assess the association between self-regulation BCTs and user ratings  
*(this analysis was added post-hoc and was not in the protocol. The rationale for including this analysis is presented in the reflective log section below)*

Objective 2:

- To assess the association between user ratings and:
  - user involvement
  - presence of any peer-reviewed studies associated with the app
  - app features
  - app usability
  - app ranking
  
- To assess the association between BCTs and
  - expert involvement
  - presence of any peer-reviewed studies associated with the app
  - app usability
  - app ranking
  - app size
  - features and BCT

As only 1 app had evidence of affiliation with a reputable organisation and there was no evidence for user involvement in the app development processes, these objectives could not be conducted.

### **The operationalisation of likely efficacy**

Initially, the definition of the likely efficacy of apps was operationalised as the number of BCTs only. The analysis assessing the relationship between the number of BCTs and other variables was conducted in other studies as discussed in the Introduction to this chapter. I provided a rationale for such operationalisation of the likely efficacy measure, however, the reductionist approach of this study was a concern for me.

The BCT Taxonomy represents the variety of smallest, irreducible, active components that are used in behaviour change interventions [168]. However, there is no evidence that more BCTs equals higher efficacy. There is evidence that inclusion of behaviour change theory might be associated with efficacy [140].

Hence, I considered an additional approach to define the likely efficacy of apps considering the opportunities (ready available data from the app stores) and limitation of this study (crude measures). I decided that I should add an analysis of those BCTs that have been shown to be effective in increasing PA to objective 1.

Systematic reviews and meta-analyses have shown that self-regulation techniques, such as goal setting, monitoring, and feedback are effective in increasing PA behaviour [140, 141, 188, 207] and that they may have cumulative effects. For example, Michie, Abraham [140] showed that self-monitoring with other self-regulatory BCTs was more effective in increasing PA than using one of those BCTs in isolation. Self-regulation has been acknowledged as an important constructs in behaviour change theories, e.g. in Control Theory [208] and Self-Regulation Theory [209]. Hence, the presence of these BCTs can be used as an indicator for quality of those apps and a proxy measure of their likely efficacy.



I found one study that looked at the individual presence of BCTs that were effective in weight management interventions. Bardus, van Beurden [136] looked at the correlation between the BCTs that were associated with greater effectiveness. They conducted correlation analysis between inclusion of self-monitoring, goal setting, and feedback and found no relationships between inclusion of these and user rating. I decided to add the analysis of those BCTs that have been shown to be effective in increasing PA behaviour. This was not an a priori analysis.

Both definitions, the number of BCTs and inclusion of BCTs that have been shown to be effective, have their weaknesses. There is no definite evidence that the more BCTs the better and there is no evidence that self-regulation techniques “work” equally in apps as they do in more traditional face-to-face interventions. However, considering the context of the study with its limited measures derived from the app stores, I thought that adding the analysis of the presence of self-regulation techniques was appropriate and needed.

\_\_\_\_\_End of the reflective log\_\_\_\_\_

### **3.3 Methods**

#### **3.3.1 Study design**

The study used regression models to determine the association between the variables derived from the app store in the sample of randomly selected popular PA apps.

#### **3.3.2 Sample identification**

The sample assessed in Study 1 was used to conduct the analysis in this study (see Chapter 2, section ‘Sample identification’). One app, Yoga Break, was excluded from this study as the data on user ratings was not available (due to unreported number of ratings), resulting in total sample of 64 apps.

### **3.3.3 Dependent Variables**

The dependent variables were the number of BCTs; provision of privacy policy; quality of development process and evidence for evaluation; the presence of expert involvement, user involvement, peer-reviewed studies; app features; SUS score. Other characteristics of the apps derived from the app stores included in the analysis were app size, app cost, ranking in the app stores.

The description of the measures of quality applied to the apps is described in details in Chapter 2, section 'Sample assessment'.

### **3.3.4 Outcome variables**

The primary outcome, user rating in the app store, was used to represent operationalisation of the user experience of the apps. In both app stores the users can award 1-5 stars to an app. The average star rating was calculated by summing the number of stars and dividing them by the number of users that submitted the rating.

When an app appeared in both stores (within the 400 apps used in identification process, see Chapter 2, section 'Sample identification'), a weighted average of the star ratings for each store was calculated based on the relative proportion of the ratings in each store. This algorithm is equivalent to summing the number of stars awarded by all users in both stores and then dividing by the total number of reviews across both stores. This calculation assumes that users are equally important regardless of the platform used to access the app. Secondary outcome used was the number of BCTs.

### **3.3.5 Identifying potential confounders**

Directed Acyclic Graphs (DAG) are visual tools used to help identify and map the possible variables that may influence the primary association of interest, and to plan which of these variables need to be adjusted for in the analysis to minimise the potential

bias in the estimation of the effect of the association of interest, in this case the BCTs and user ratings. [210, 211]. First, the potential variables that can potentially relate to both the predictor and the outcome variable were identified based on the literature review of the previous studies that assessed the association between the app content and characteristics (Figures 19 A and B) and the quality indicators and app characteristics identified in Study 1 (Chapter 2). Next, The DAG was then used to identify the potential confounders – the variables that are associated with the predictor and outcome of interest but not in the causal pathway. The process summarising the assumptions about the association of interest using DAG is found in the Appendix G. Variables determined a priori potentially to affect the relationship between BCTs and star ratings were the number of features, whether the app was free or required payment (cost), size (in megabytes, MB) and usability.

### **3.3.6 Sample size justification**

The sample size calculation was based on pilot sample of 10 apps (5 paid from iTunes, 5 free from GP) selected from the sample identification of the 100 top ranked apps in each store. Apps were sorted in order of store rankings. From iTunes, 38 potentially eligible paid apps were identified and every 8<sup>th</sup> app was included in the pilot sample ( $N = 38/5 = 7.6 = \sim 8$ ). From GP, 55 potentially eligible free apps were identified and every 11<sup>th</sup> app was included in the pilot sample ( $N = 55/5 = 11$ ). If a sample app was downloaded and found to be ineligible, the next lowest ranked app was used instead. Three apps from GP and none from iTunes were found to be ineligible and replaced.

The histogram of the star rating distribution is displayed in Figure 20. The skew that was apparent in the larger sample was not pronounced in the pilot data. The pilot data were used to estimate the standard deviations (SD) of BCTs and star ratings. Linear regression was also performed to estimate the relationship between BCTs and star ratings. The results were SD of 2.8 and 0.77 for BCT and star ratings, respectively. The regression coefficient estimate for the effect of BCTs on star ratings was 0.11 (SE=0.09,

$R^2 = 0.159$ ). This result suggested that each additional BCT unit would lead to an average increase in the star rating of 0.11. This coefficient was not statistically significantly different from 0 ( $p = 0.254$ ), but this is expected as this pilot study was underpowered to see a significant result ( $N = 10$ ). Figure 21 presents the scatter plot of the data and the linear regression line.

Figure 20: The histogram of the star rating distribution

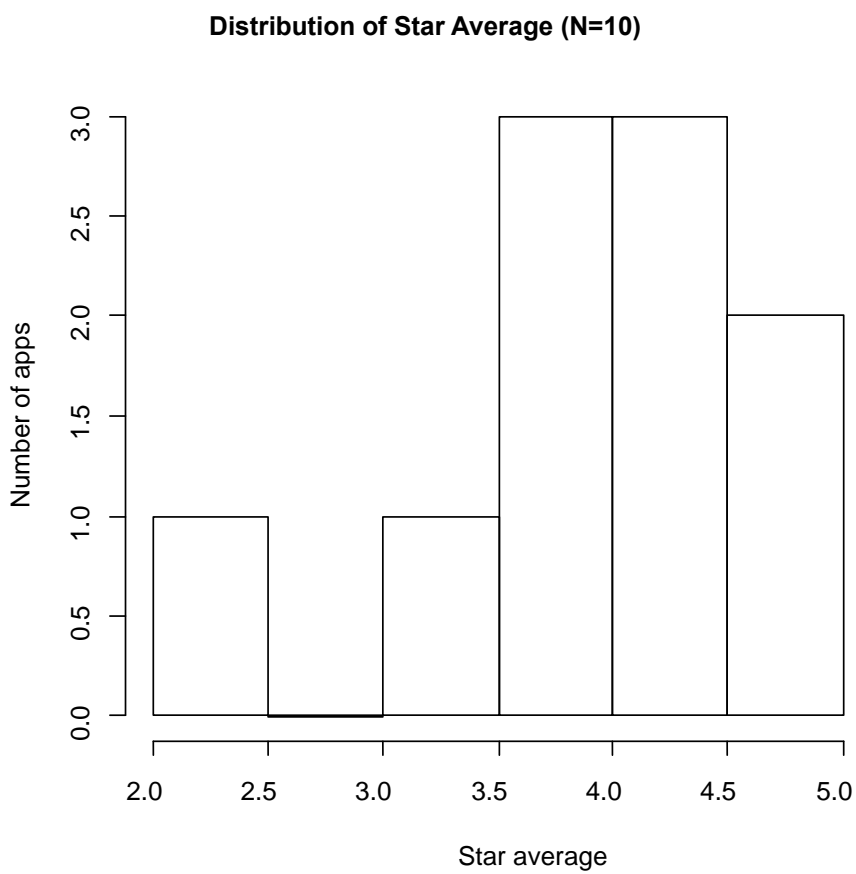
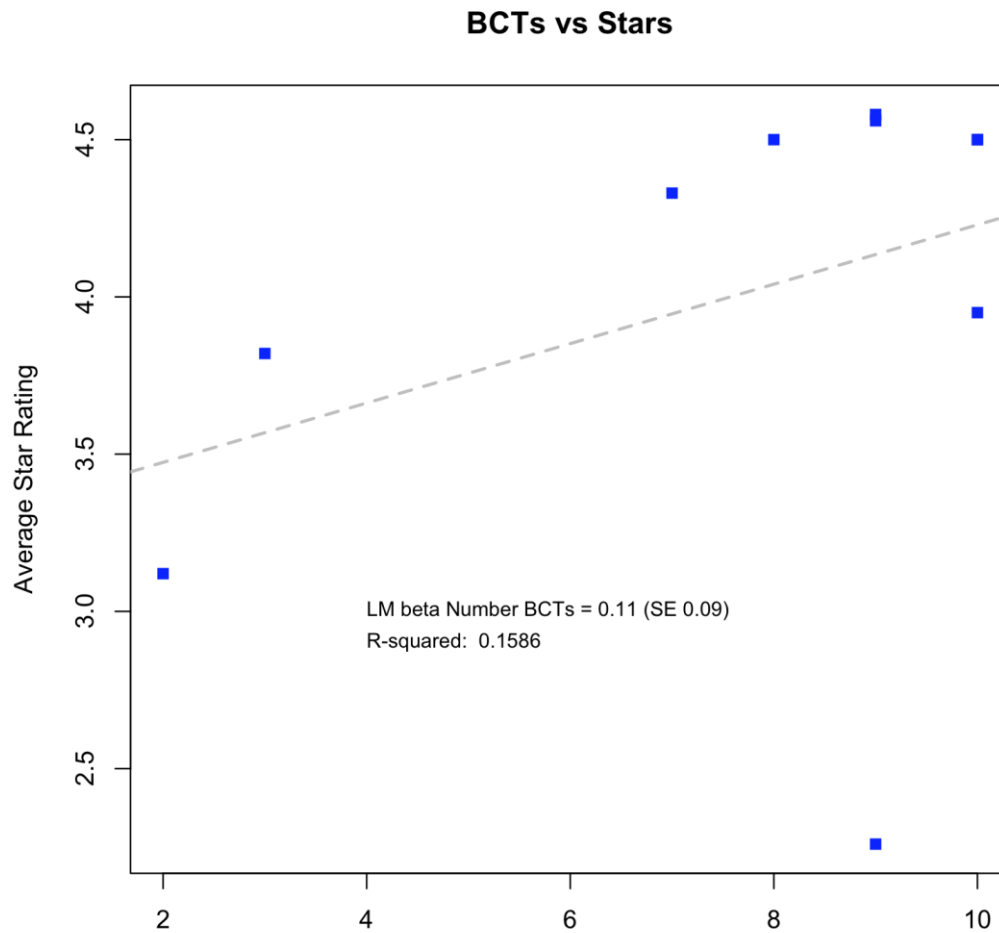


Figure 21: Scatter plot with linear regression of the pilot data (N=10)



The sample size was then calculated assuming a relatively symmetric distribution of star ratings [212] (see Appendix H for sample size calculations). Statistical power of at least 80% is recommended and 90% is desirable. Based on these calculations, a sample size of 51 apps would provide 90% power to detect a change of 0.11 for each additional BCT at 5% significance level (type I error rate). A sample size of 65 apps was selected to account for any randomly selected apps that might not fulfil the inclusion criteria once downloaded. If all 65 apps are eligible, the resulting study would have 94% power to detect a change a mean change of 0.11. These calculations assume that pilot data are similar to the study data in that the mean change is 0.11. If the relationship was smaller than 0.11, calculations showed that a sample size of 65 would still provide 80% power to detect a mean change in star rating of 0.087 for each additional BCT at 5% significance

level. A mean change of 0.087 is approximately a 20% reduction from the 0.11 found in the pilot data, and any smaller change suggests that the pilot data are not a representative sample of possible apps.

### **3.3.7 Statistical analyses**

The number of BCTs in the apps was summarized using the mean, standard deviation, median, 25<sup>th</sup> and 75<sup>th</sup> percentiles, and the maximum and minimum. Similar statistics was used to summarize user ratings, cost, as well data privacy and security assessment, and SUS.

T-tests were used to assess the difference between app characteristics and the number of BCTs. Specific tests were undertaken to compare the difference in the number of BCTs between apps with and without the evidence of user involvement, with and without the evidence of the presence of peer-reviewed studies, apps with higher and lower usability, and larger and smaller apps.

Three types of regression models were used to analyse the study data. In light of the strengths and limitations described below, the primary analysis is based on a linear regression model and powered accordingly. Logistic and proportional odds models will be calculated to examine some of the limitations of the linear model. Analyses were performed using SAS version 9.3 and R version 3.3.3.

#### **Linear regression**

Linear regression investigates the relationship between 2 continuous variables. It is used to determine how one variable (the predictor of interest) relates to another variable (the outcome of interest). The result of a regression model is the intercept and slope of the best fitting straight line through the data. This information can be used to predict how

values or changes in the predictor of interest might impact the outcome variable. Linear regression also allows for control of other variables (covariates). These additional variables may be used to improve the estimate for the predictor of interest, or to control for potential confounders.

The key assumptions of linear regression are:

1. The dependent variable needs to be continuous.
2. *Linearity* - As the statistical analysis involves fitting a straight line onto the data, the relationship between the outcome and predictor needs to be linear. The linearity assumption can best be checked by inspecting the scatter plots. For example, if there is a pattern (e.g., a curve) then the assumption of linearity may be violated. In addition, it is important to check for outliers and influential cases which can affect the model.
3. *Independence of errors* (also described as lack of autocorrelation) - the residuals must not be correlated. When observations are independent, errors are independent. In this study, apps are assumed to be independent so errors are independent.
4. *Homoscedasticity* (homogeneity of variance) - that the error terms at each level of predictor variable are equal, i.e., they have the same variance. Diagnostic plots should be obtained to investigate for heteroscedasticity in the data.
5. *Normally distributed errors* - the residuals should be normally distributed with a mean of 0. This means that the difference between the model and the data is assumed to be 0 or close to 0. The normality of residuals assumption can be investigated using histograms and P-P plots. Deviations from a bell-shaped curve and non-diagonal line in P-P plots may suggest non-normally distributed residuals. Diagnostic plots should be obtained to investigate for non-normally distributed errors.
6. *Multicollinearity* – It occurs where there are strong correlations ( $r > .8$ ) between explanatory variables and the outcome. It increases the standard error of the regression

coefficient which means they are less likely to represent the population, and it decreases the value of  $R^2$ . This assumption can be checked by inspecting a correlation matrix for all included variables.

#### Advantages:

Linear regression is a method which is commonly used to assess the relationship between variables. It has also been used in the context of evaluating apps [137], so there is precedent in the literature for this technique.

Linear regression does not require dividing star ratings into categories. The average star rating can be used as the outcome variable. Also, the interpretation of the regression is appropriate for the study as it shows the average effect of BCTs on star ratings.

#### Disadvantages:

It assumes that the distance between any two stars is the same as between other stars, i.e., any additional BCT would improve 3-star rated app exactly the same amount as it would improve a 1-star rated app.

Linear regression performs well for symmetric variables without major outliers. Star average is bounded between 1 and 5, though it is likely to be skewed toward high values. The distribution of the number of BCTs may be markedly skewed if a few apps have a lot of BCTs, though the mean and median are expected to be approximately 6. We can attempt to resolve these issues with log transformations.

The relationship between stars and BCTs may not be linear – the mean impact of additional BCT units may be large when there are only a few, but not as large when there are many BCTs already included. Transformations of both the predictor and the outcome may be used to improve model fit.



## Logistic regression

In the present study, star ratings will be represented as a binary variable: *higher user ratings* of 4 or 5 stars versus *lower user rating* of 1, 2, 3 stars. The interpretation for the odds ratios for the BCTs will be the increase in odds of higher user rating (stars 4 or 5) associated with each BCT element.

Logistic regression is similar to linear regression in that it quantifies the relationship between predictors and outcomes but in this case, the outcome is a binary variable. Results are interpreted on a probability scale and the most commonly reported statistic is the odds ratio.

As logistic regression is a generalized linear model, the assumptions of this model are similar to that of linear regression. However, because in this case the dependent variable is categorical with only 2 levels, the *assumption of linearity* (predictor has linear association with the outcome) is not relevant. Instead, log (or logit) of the outcome variable is used and the explanatory variable needs to be associated with the log of the outcome.

### Advantages

The interpretation of the odds ratio is related to the probability of awarding a higher star rating as opposed to a lower star rating, which is more intuitive than an average increase in star average (which is the interpretation of the linear regression coefficient).

### Disadvantages

The choice of threshold (4 and 5 versus 1, 2 and 3 stars) is somewhat arbitrary. Sensitivity analyses can be used to determine how much the results depend on the choice of threshold – for example if 4 stars was classified with lower ratings or 3 stars was classified with higher ones.

The logistic model results in loss of granularity, in comparison to a continuous model. While users get to select between 5 stars to rate the app, logistic regression reduces the choice to 2 categories: 1, 2, 3 stars versus 4 or 5 stars. We cannot quantify the effect of BCTs on, for example, a 1-star increase using linear regression.

### **Proportional odds model**

The proportional odds model is similar to logistic regression in that the outcome is categorical, but there can be multiple categories instead of just 2. Logistic regression models can be performed for all choices of thresholds that could be used in logistic regression: 1 versus 2, 3, 4, 5, and 1, 2 versus 3, 4, 5, and 1, 2, 3 versus 4, 5 and 1, 2, 3, 4 versus 5. This is essentially what the proportional odds model does, but it combines these 4 models into a single model and so makes better use of all the information in the data.

#### Advantages

The analysis is conducted on what the users reported, i.e., in this model, collapsing the outcome based on a threshold is not necessary.

This model does not assume equal distance between stars, as is the case for linear regression. The proportional odds model only requires that the categories are ordered, which is the case for star ratings, i.e., more stars are better. Furthermore, the distance between stars need not be consistent from person to person. The model can also accommodate differences in the interpretations across stars – one person may award 5 stars to any app that works as described, while another would only award 5 stars to an app that exceeds the description.

This model also addresses the primary goal of the study, as the interpretation of the odds ratio is the impact of additional BCT units on awarding a higher star rating as opposed to impacts on the average star rating or the chance of awarding 4 or 5 stars.

## Disadvantages

The main disadvantage of this model is that it relies on the proportional odds assumption and cannot be calculated if this assumption is not met.

In this study, the proportional odds assumption implies that an additional BCT increases the chance of a higher star rating by the same amount regardless of the starting star level, i.e., the BCT increases the chances of going from star 4 to 5 by the same amount as going from star 1 to 2. In reality, it is possible that BCTs may make a poor quality app into good apps (improve apps that are of low quality) but once the app has 4-star rating, an additional BCT unit might not have as much influence. In this case, the proportional odds assumption is likely to be violated.

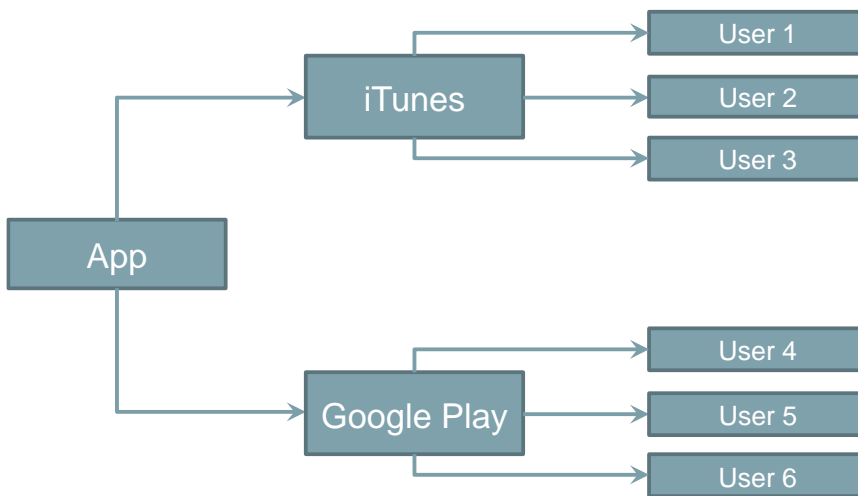
Proportional odds models are a relatively recent invention, and their use in user feedback situations is sparse. The model has been used in the context of user feedback, i.e., patient reported outcome but has not been applied to BCTs [213].

It is the model of choice for a categorical response as it does not assume linearity and it does not collapse categories. However, it might not work because the proportional odds assumption may fail.

### 3.3.7.1.1 Data clustering

Simple regression models assume the independence of records. For this study, this assumption would mean that around 2.8 million users rated 2.8 million apps. In fact, the data in the study includes 2.8 million users ratings 65 apps from 2 different stores. (see Figure 22 for the data structure in the study). Hence, appropriate adjustments were made to the model using fixed effects for the app stores, and random effects for the apps.

Figure 22: Data clustering in the current study



An effect in a study is **fixed** when all conditions are present in the experiment. Fixed effects can only be generalised to the situations similar to those in the study [214]. In the current study, fixed effects were used to adjust the data for the app store as the majority of apps in the app store come from these 2 app distribution platforms. The interpretation of the results using fixed effects holds *for these stores* (and not other potential app distribution platforms).

An effect in a study is **random** when the conditions present in the experiment are random sample of the possible conditions. Random effects can be generalised to the situations beyond the conditions in the experiment [214]. With random effects, however, the CIs are larger, hence there is less certainty in the results. In this study, random effects were used to adjust the data for the PA apps as 65 apps were randomly selected from the total number of 125 apps in the top rankings of the apps stores. Random effect assume that different apps could have been randomly selected which is true for the current data as the focus of the study lies in the underlying PA app population. The interpretation of the results using fixed effects holds *for popular PA apps in general*.

### 3.3.7.1.2 Sensitivity analysis plan

Sensitivity analysis was performed to explore whether the overall results of the analysis were affected by the type of apps included, the threshold used to dichotomise the outcome variable, the low number of user ratings, and the confidence level in the presence of the BCT extracted. More specifically, sensitivity analysis that will be conducted is listed:

1. Including only those BCTs that were classified as “present beyond all reasonable doubt” (++)
2. Analysis of those apps that are appropriate for beginners
3. Changing the threshold for outcome (star ratings – 5 stars vs 1,2,3,4 )
4. Excluding apps with less than 25 user ratings

## 3.4 Results

### ***3.4.1 Descriptive statistics for the independent variable***

The mean for the number of BCTs variable was 7.1 (SD= 3.0) and was normally distributed. The details for the presence of BCTs in apps are described in Chapter 2, section: ‘The presence of BCTs’ (Table 20):

Table 20: Descriptive statistics for the predictor: number of BCTs

		<b>Free (N=32)</b>	<b>Paid (N=32)</b>	<b>Total (N=64)</b>
Number of BCTs	Mean $\pm$ SD	6.56 $\pm$ 2.99	7.56 $\pm$ 2.87	7.06 $\pm$ 2.95
	Median	7.00	8.00	8.00
	25, 75%tile	5.00, 8.00	6.00, 10.00	5.00, 9.00
	Min, Max	1.0, 12.0	1.0, 13.0	1.0, 13.0

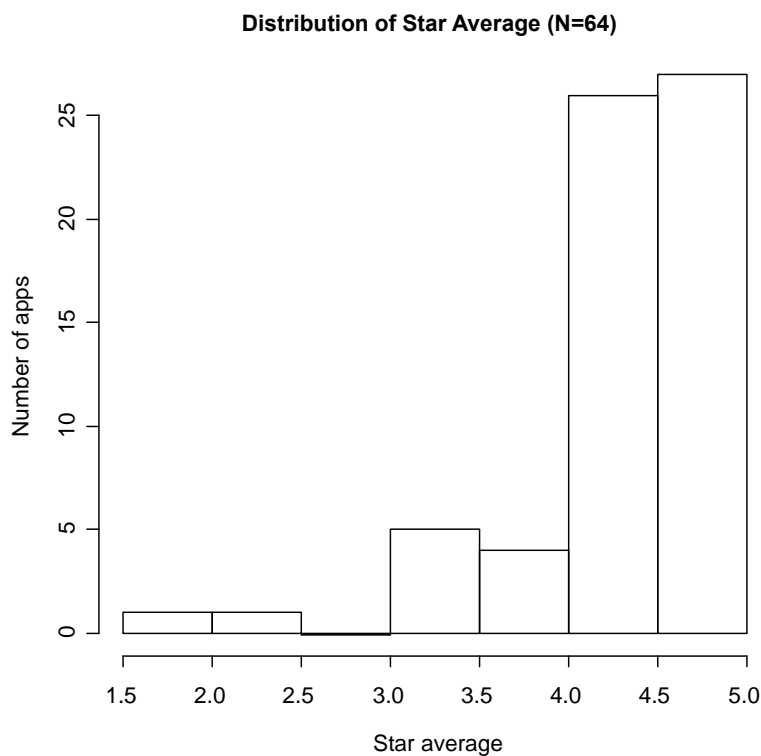
### **3.4.2 The distribution of the outcome variable**

In total, there were 2,819,469 individual user ratings from 64 apps used in the study (89,373 in iTunes and 2,730,342 in GP). Among these, 88.5% were 4 or 5 star reviews and 11.5% were 1, 2 or 3-star reviews. See Table 21 for the descriptive statistics of the star ratings for both stores. The distribution of star average was highly skewed (coefficient of skew -2.48, see Figure 23 for the distribution of star average for both stores). The acceptable values for skew are  $\pm 1$  [215].

Table 21: Descriptive statistics for the weighted star average of both stores (app level data)

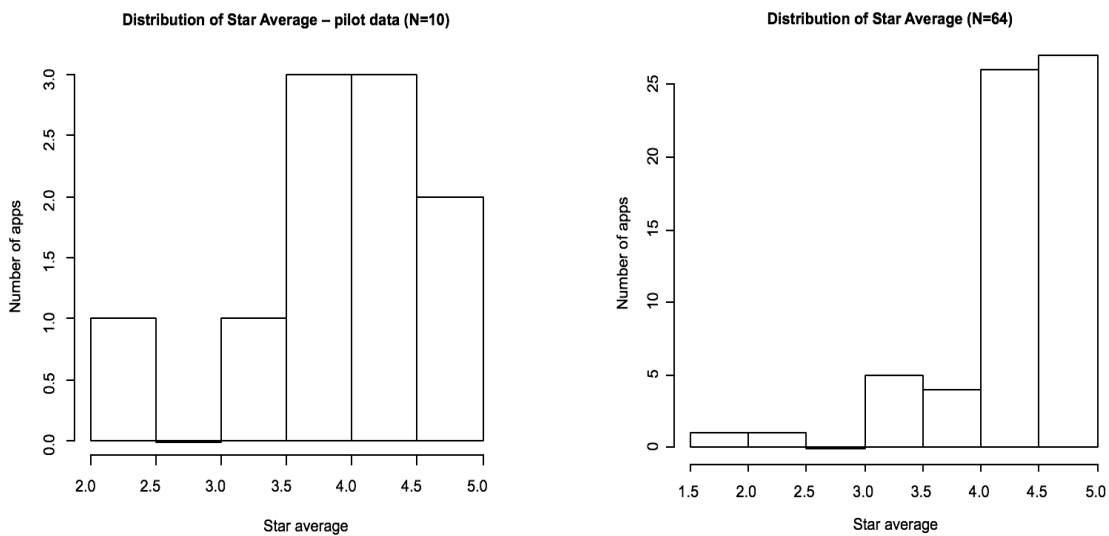
		<b>Total - iTunes (N=44)</b>	<b>Total - GP (N=37)</b>
User rating (1-5 stars)	Mean $\pm$ SD	4.20 $\pm$ 0.71	4.36 $\pm$ 0.42
	Median	4.40	4.50
	25, 75%tile	4.00, 4.60	4.40, 4.60
	Min, Max	1.8, 4.9	2.4, 4.7
Number of ratings	Mean $\pm$ SD	2031 $\pm$ 4290	73793 $\pm$ 137000
	Median	550.0	5856
	25, 75%tile	85.50, 1719	1475, 78204
	Min, Max	11.0, 25000	7.0, 630000

Figure 23: Distribution of star average for both stores



The pilot data failed to show the skew of the outcome variable and log transformation of the data were not possible due to the direction of the skew (see Figure24) below comparing the star average distribution between the pilot data (N= 10) and full sample (N=64).

Figure 24: star average distribution between the pilot data (N= 10) and full sample (N=64)



Consequently, linear regression (as per the analysis plan) was not conducted as the degree of skew for the outcome variable would likely lead to violation of the normally distributed errors assumption. While the logistic regression analysis was pre-specified and conducted as planned (section ‘Statistical analyses’), it is described in the analysis plan as the secondary analysis which required representing star ratings as a dichotomous variable. Proportional odds modelling was also not conducted due to the skew of the outcome variable.

The outcome variable was dichotomised into higher user ratings of 4 or 5 stars versus lower user rating of 1, 2, 3 stars.



### **3.4.3 Descriptive statistics for the independent variable and confounders analysis**

A set of variables identified in the analysis plan for the relationship to star rating (cost, usability, number of features, size, see section 'Identifying potential confounders') were included as fixed covariates in univariate logistic regression models with the outcome of high versus low star ratings. Cost, size, and usability were skewed and were dichotomised. As there is no guidance on what constitutes a "large" or "small" app, apps above the median size value (82.2 MB in iTunes and 29.6 MB in GP) were classified as large apps, and those below were classified as small. For cost, apps were either free or paid. For the SUS score a threshold validated by Bangor, Kortum [165] was used with 72.5 described as good usability.

Descriptive statistics for the confounder variables are displayed in Table 22. The mean number of features was 6.3 (SD=2.7). Among paid apps, the median cost was £2.4 (IQR 1.78 – 2.99). The SUS score and the cost variables were skewed. The median size of apps was 39.64 MB (IQR 25.52– 79.50) and the median SUS score was 86.3 (IQR 75.00 – 91.88).

Table 22: Descriptive statistics for the analysis of the confounders

		<b>Free (N=32)</b>	<b>Paid (N=32)</b>	<b>Total (N=64)</b>
Number of Features	Mean $\pm$ SD	5.94 $\pm$ 2.17	5.72 $\pm$ 1.95	5.8 $\pm$ 2.05
	Median	7.00	8.00	7.00
	25, 75%tile	4.00, 8.00	4.50, 7.00	4.00, 8.00
	Min, Max	1.0, 8.0	1.0, 9.0	1.0, 9.0
Cost (GBP)	Mean $\pm$ SD		3.00 $\pm$ 2.12	
	Median	N/A	2.35	N/A
	25, 75%tile		1.78, 2.99	
	Min, Max		0.8, 9.0	
Size (MB)	Mean $\pm$ SD	45.53 $\pm$ 40.85	61.51 $\pm$ 47.22	53.52 $\pm$ 44.53
	Median	32.45	53.97	39.64
	25, 75%tile	16.82, 62.27	28.92, 85.02	25.52, 79.50
	Min, Max	1.9, 163	1.0, 242	1.0, 242
Usability (SUS)	Mean $\pm$ SD	81.25 $\pm$ 12.64	85.23 $\pm$ 12.02	83.24 $\pm$ 12.40
	Median	85.00	87.50	86.25
	25, 75%tile	71.88, 91.25	78.75, 93.75	75.00, 91.88
	Min, Max	52.5, 100	57.5, 100	52.5, 100

Note: MB: megabytes, SUS, System Usability Scale; GBP: British pounds, N/A, not applicable

Any variables significant at  $p \leq 0.1$  were included in models examining BCTs. Usability was associated with odds of assigning higher user ratings and was therefore controlled for in the model (see Table 23).

Table 23: The analysis using univariate logistic regression models with the outcome of high versus low ratings.

<b>Variable</b>	<b>Odds ratio</b>	<b>Lower 95% CI</b>	<b>Upper 95% CI</b>	<b>p-value</b>
Number of features	1.10	0.97	1.24	0.134
Cost (free vs paid)	0.78	0.48	1.28	0.329
Size (large vs small)	0.98	0.94	1.02	0.221
Usability (high vs low)	1.66	0.96	2.89	0.071

### **3.4.4 Primary Analysis of objective 1**

#### **3.4.4.1 Assessment of the association between the number of BCTs and user ratings.**

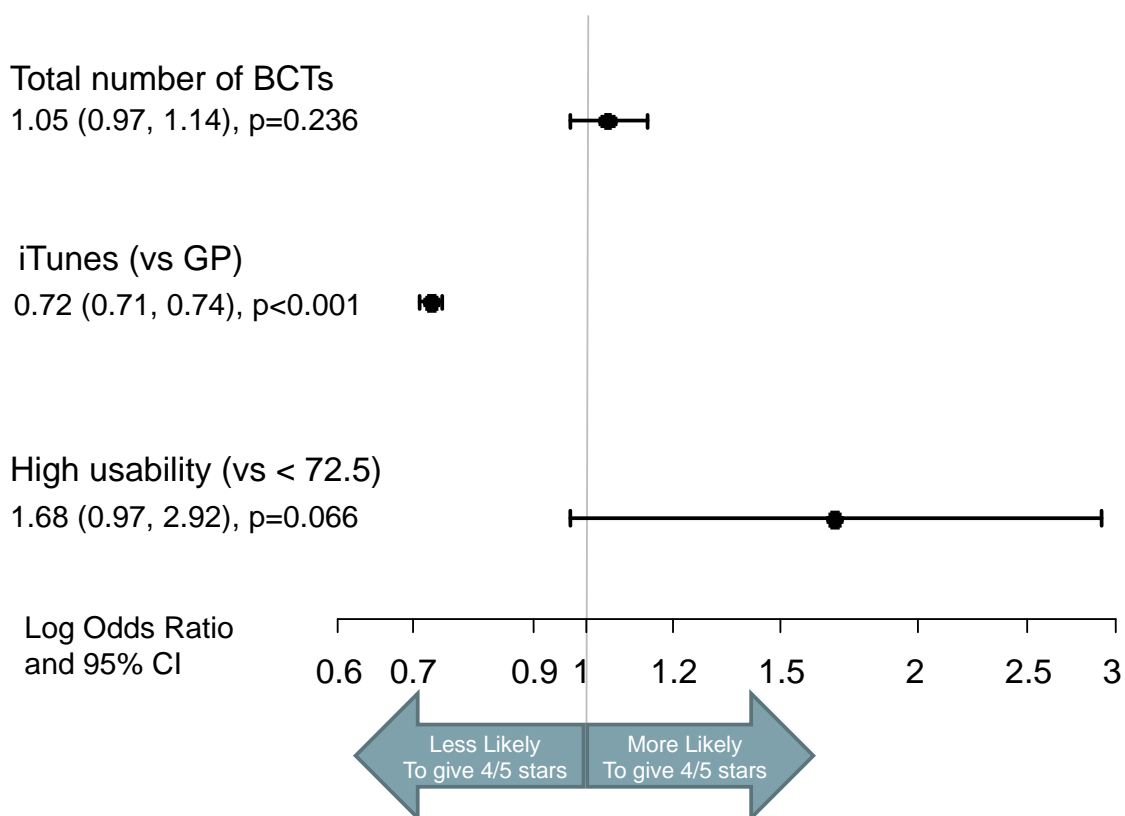
Weighted logistic regression (by the number of responses for each app) was used with a random intercept term for app, reflecting that the apps selected for analysis were a random sample of available fitness apps, and to control for correlation when the same app was reviewed in both stores.

See Figure 25 for the results of the primary analysis. There was no significant association between the number of BCTs and the likelihood of assigning higher user ratings (OR = 1.05, CI 95% 0.97 to 1.14,  $p = 0.236$ ).

The association between the store and star average was significant. Among users evaluating the PA apps with the same number of BCTs and the same usability (high or low), being in iTunes is associated with 37% reduction in the odds of a 4 or 5 star rating (OR = 0.72, CI 95% 0.71 to 0.74,  $p < 0.001$ ).

The association between the usability and star average did not reach statistical significance (OR = 1.68, CI 95% 0.97 to 2.92,  $p = 0.066$ ).

Figure 25: The graph showing the results of the primary analysis



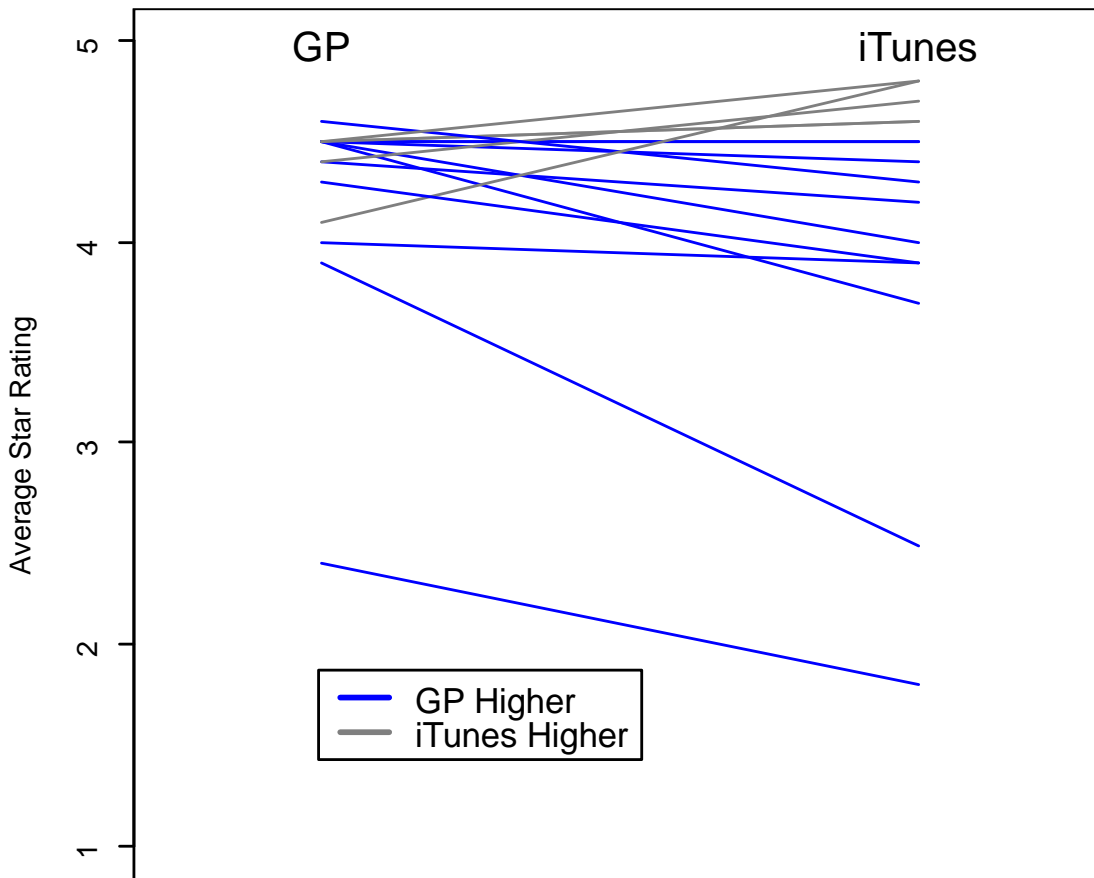
### Users in iTunes and GP differ in how they assign star ratings

The significant result of the primary analysis suggesting that users behave differently in how they assign user ratings depending on store (iTunes or GP) was unexpected. The assumption of the study was that users behave similarly in terms of the decision to assign 1 to 5 stars regardless of the store. This assumption therefore had to be revisited and the association between the app stores and user ratings were explored further.

First, the star ratings of the duplicate apps, i.e., the same apps available in both stores (n= 17), were investigated. Figure 26 visualises the difference between the star ratings in iTunes versus GP. Each line represents 1 mobile app. The blue line indicates when, for the same app, star ratings were higher in GP. The grey lines show when the ratings were higher in iTunes. There were only 4/17 cases when star ratings were higher in iTunes than in GP for the same app.

Analysis was conducted to assess whether the proportion with the high star ratings in GP exceeds what would have been found by chance. If there was no relationship between store and star rating, we would expect the star rating in iTunes to exceed the star rating in GP for about 50% of apps (coin flip). Mean star rating in GP exceeded the rating in iTunes in 13/17 apps (76.5%, 95% CI: 64.2-86.2,  $p < 0.0001$ ). Since the analysis was significant, we infer that the difference between the user ratings in the stores was not due to chance, and that GP users were more likely to assign high ratings than iTunes users.

Figure 26: Star ratings for the same apps across the app stores (n= 17)



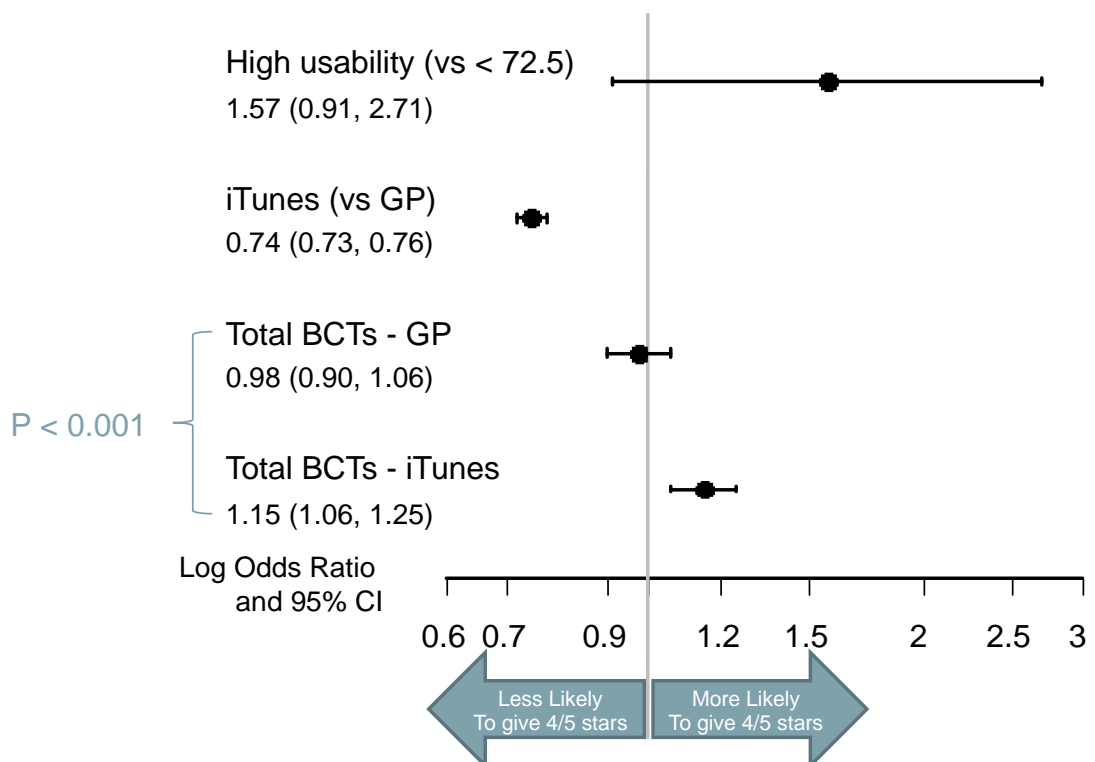
### The effect of total number of BCTs is not consistent across stores

Following the finding that users differ in how they assign star ratings depending on the store, an interaction analysis between the store and the number of BCTs was conducted. Figure 27 shows the primary model with an interaction between the store and the number of BCTs. The interaction between the store was significant which indicates that the effect of BCTs differs across the app stores. There is no effect for additional BCTs in GP (OR= 0.98, CI 95% 0.90 to 1.06) but there is a significant effect for iTunes (OR= 1.15, CI 95% 1.06 to 1.25). Specifically, each additional BCT increases the chances of higher user rating in iTunes by 15%.

To account for the significant interaction, which indicated that the effect of the BCTs differs across the app stores, the store was added to the model as an interaction. This

was specific to the analysis where the outcome was user ratings (objective 2). This analysis was not planned a priori.

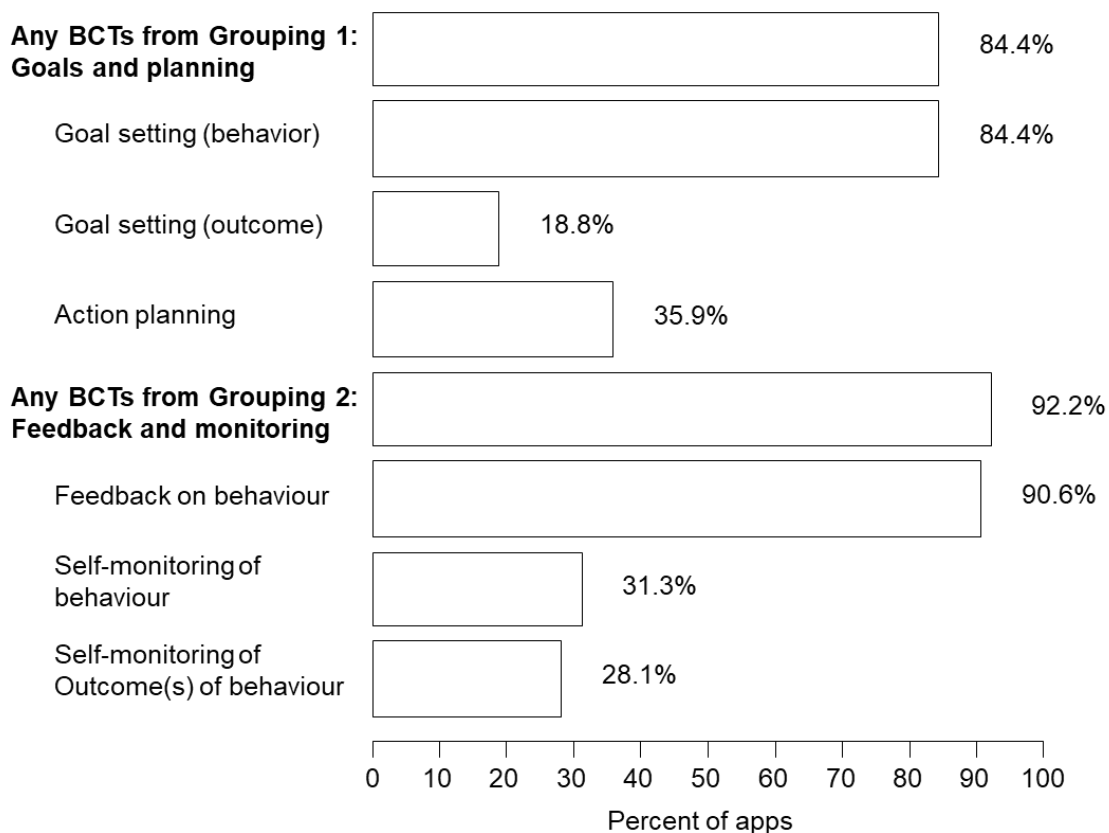
Figure 27: The primary analysis of the number of BCTs with user ratings with the addition of an interaction between the store and the number of BCTs



### 3.4.4.2 Assessment of the association between self-regulation BCTs and user ratings

The BCT grouping 1 (Goals and planning) was represented in 84.4% of evaluated apps. The most common of these was goal setting for behaviours (84.4%), followed by action planning (35.9%) and goal setting for outcomes (18.8%). The BCT grouping 2 (Feedback and monitoring) was included in 92.2% of evaluated apps. Feedback on behaviour was included in nearly all apps (90.6%), while self-monitoring of behaviour and of outcomes were incorporated in about a third of apps.

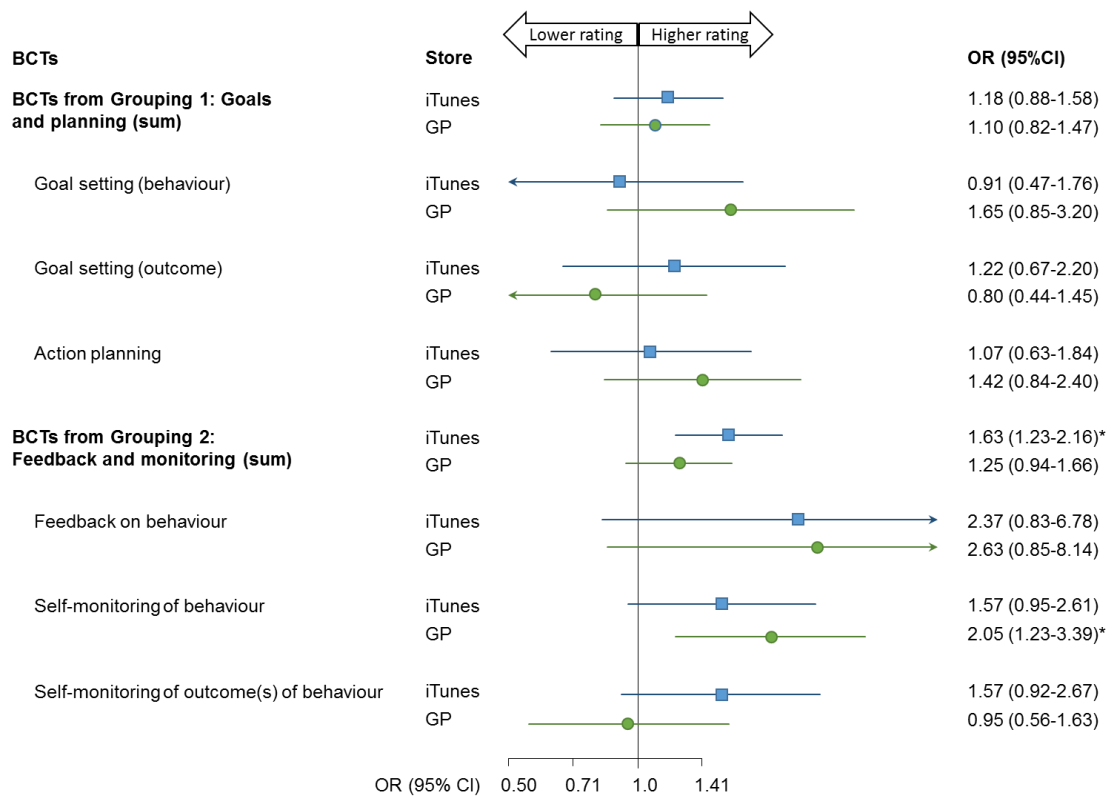
Figure 28: Frequency of individual BCTs within the two Groupings of self-regulatory BCTs



The number of Grouping 1 BCTs had no impact on star rating in either store (Figure 29). Among iTunes users, each additional BCT from Grouping 2 increased the chance of a high star rating by 63% (OR: 1.63 95% CI: 1.23-2.16). There was no association between the number of BCTs and star rating among GP users. Self-monitoring of behaviour was associated with higher ratings in GP (OR: 2.05, 95% CI: 1.23-3.39)



Figure 29: Odds ratios for the associations between BCTs that have been shown to be effective and a 4- or 5- star rating



### 3.4.5 Secondary analysis of objective 2

### 3.4.6 User ratings as outcome

A series of logistic regressions were conducted to assess the association between the variables and the user rating outcome.

#### 1. Is there a relationship between expert involvement and star average?

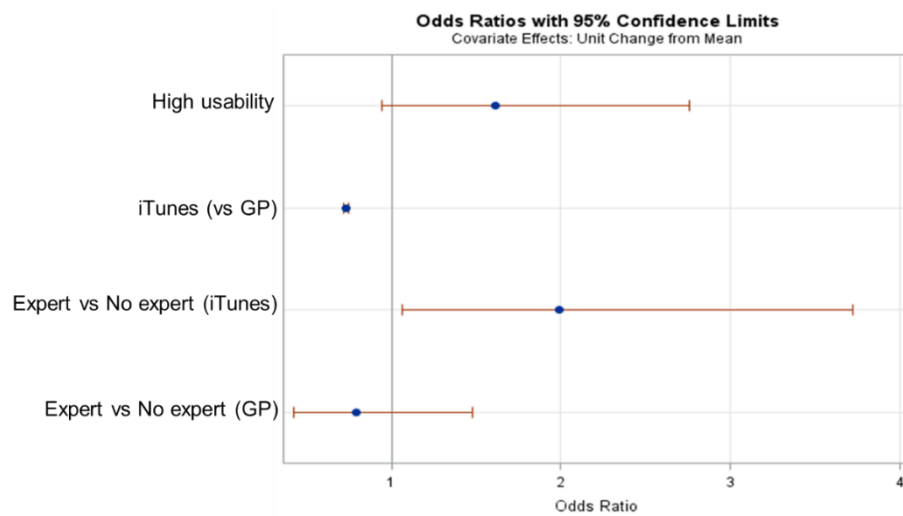
Out of 64 apps included in the analysis, 12 apps reported involving experts in the development of the app. The analysis using logistic regression showed a significant interaction between expert involvement and app store which suggests that the effect of expert involvement on star average is different across the app stores ( $p < .0001$ ).

Specifically, there was no association between expert involvement and star ratings in GP (OR: 0.79, 95% CI: 0.42-1.48), however, there was an association in iTunes (OR: 1.99, 95% CI: 1.07-3.72) with apps with expert involvement being 1.99 times more likely to receive higher user ratings (4 or 5 stars) than those that did not involve experts (Table 24).

Table 24: Odds ratios for the association between expert involvement and star average with expert involvement and app store interaction

<b>Expert involvement and star average</b>	<b>OR</b>	<b>95% CI</b>		<b>p-value for the interaction</b>
iTunes	1.99	1.07	3.72	<.0001
GP	0.79	0.42	1.48	

Figure 30: Logistic regression showing the effect of the expert involvement on user ratings (expert involvement x store interaction)



## 2. Is there a relationship between the number of features in the app and star average?

The descriptive statistics for the number of features within apps is presented in Chapter 2, section 'App features'.

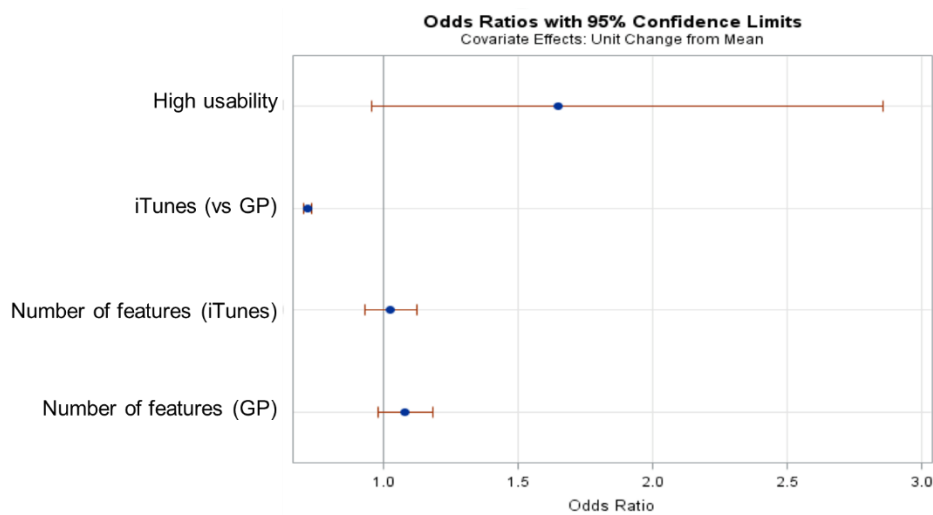
As indicated in section 'Descriptive statistics for the independent variable and confounders analysis', the distribution of the number of features in apps was relatively normal (skew -0.27, kurtosis -0.60).

There was no significant association between the number of app features and star average in either iTunes or GP (Table 25). The interaction was statistically significant ( $p < .0001$ ) but the odds ratios showed no meaningful difference.

Table 25: Odds ratios for the association between the number of app features and star average with number of app features and app store interaction

Number of features and star average	OR	95% CI	
iTunes	1.02	0.93	1.12
GP	1.08	0.98	1.18

Figure 31: Logistic regression showing the effect of the number of app features on user ratings (number of app features x store interaction)



### 3. Is there a relationship between usability and star average?

The descriptive statistics for the SUS score are presented in Chapter 2, section 'User experience'. As described in section 'Descriptive statistics for the independent variable and confounders analysis', the SUS score was skewed and was dichotomised based on the published threshold [216].

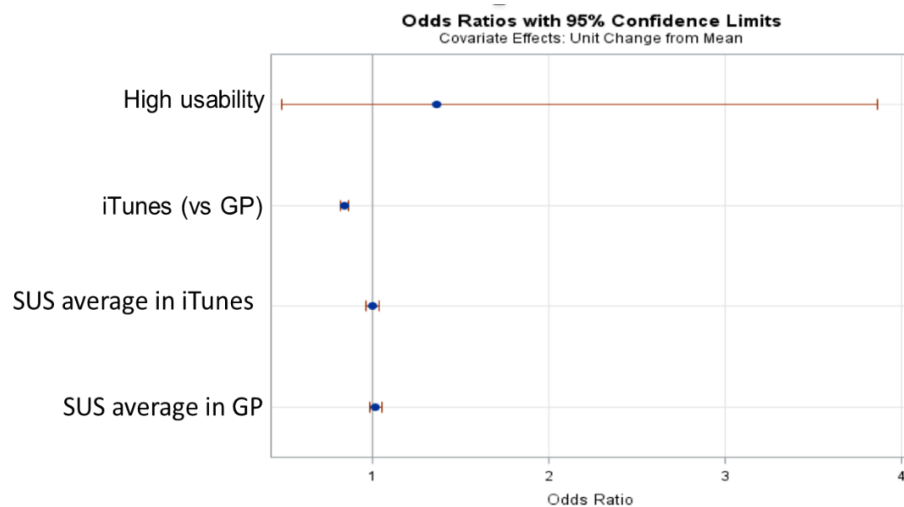
Logistic regression showed a significant interaction between usability and app store suggesting the effect of usability on star average is different across the app stores

( $p < .0001$ ). There was no association between usability and star average in either iTunes or GP (Table 26).

Table 26: Odds ratios for the association between usability and star average and app store interaction

Usability and star average	OR	95% CI		p-value for the interaction
iTunes	1.00	0.96	1.03	<.0001
GP	1.02	0.98	1.06	

Figure 32: Logistic regression showing the effect of usability on user ratings (usability x store interaction)



#### 4. Is there a relationship between app ranking and star average?

As the rankings are separate for free and paid apps, the analysis was conducted separately for free and paid apps.

### a) Free apps

Logistic regression showed a significant interaction between app store ranking for free apps and app store suggesting the relationship between the ranking and star rating differ by store ( $p < .0001$ ). There was no association between app store ranking for free apps and star average in either iTunes or GP (Table 27).

Table 27: Odds ratios for the association between app store ranking for free apps and star average with app store interaction

<b>App store ranking (free apps)</b>	<b>OR</b>	<b>95% CI</b>		<b>p-value for the interaction</b>
iTunes	0.96	0.93	1.00	<.0001
GP	0.99	0.95	1.03	

### b) Paid apps

There was also no association between app store ranking and star average among paid apps (Table 28).

Table 28: Odds ratios for the association between app store ranking for paid apps and star average with app store interaction

<b>App store ranking (paid apps)</b>	<b>OR</b>	<b>95% CI</b>		<b>p-value for the interaction</b>
iTunes	1.00	0.96	1.04	0.021
GP	0.98	0.94	1.01	

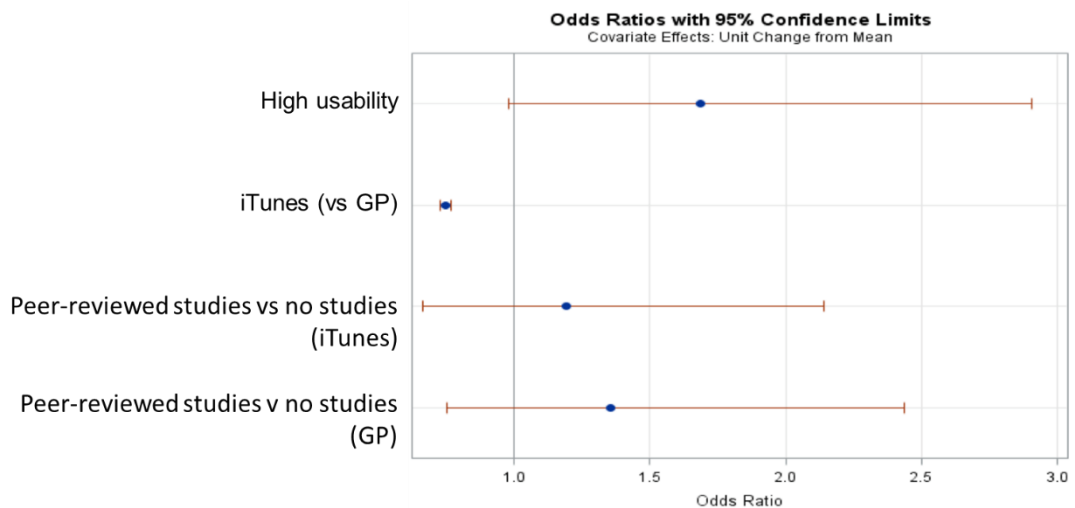
**5. Is there an association between the presence of peer-reviewed studies and star average?**

There was no association between peer-reviewed studies and star average in either iTunes or GP (Table 29). The interaction was statistically significant ( $p=.0001$ ) but the odds ratios showed no meaningful difference.

Table 29: Odds ratios for the association between the presence of peer-reviewed studies and star average with the presence of peer-reviewed studies

<b>The presence of peer-reviewed studies and star average</b>	<b>OR</b>	<b>95% CI</b>	
iTunes	1.19	0.66	2.14
GP	1.36	0.75	2.44

Figure 33: Logistic regression showing the effect of the presence of peer-reviewed studies on user ratings (the presence of peer-reviewed studies x store interaction)



### 3.4.7 BCTs as outcomes

#### 6. Is there an association between the expert involvement in development and BCTs?

The number of BCTs within app was normally distributed, so the relationship between expert involvement and the mean number of BCTs as an outcome was analysed using Student t-test.

For 12/64 apps there was evidence that experts were consulted in the development process. For apps with no evidence of expert involvement, the mean number of BCTs was 7.15. For those that involved experts, the mean BCT number was 6.5. An independent t-test showed there was no difference in the mean number of BCTs between the apps that included and did not include expert involvement (0.65, 95% CI: -1.24 to 2.54,  $p=0.493$ ), see Table 30.

Table 30: The results of the t-tests comparing the number of BCTs between apps with and without evidence of expert involvement

Expert involvement	n	Mean (SD)	95% CI		p-value
Yes	53	7.15 (2.91)			
No	12	6.50 (3.15)			
Difference	-	0.65	-1.24	2.54	0.493



## 7. Is there an association between usability and BCTs?

The descriptive statistics for the SUS score are presented in Chapter 2, section 'User experience'. As described in section 'Descriptive statistics for the independent variable and confounders analysis', the SUS score was skewed and was dichotomised.

For apps with high usability, the mean BCT number was 6.98. For apps with lower usability the mean number of BCTs was 7.2. There was no statistical difference between mean number of BCTs between tabs with high versus low visibility (-0.22, 95% CI -1.96 to 1.52,  $p=0.802$ ), see Table 31.

Table 31: The results of the t-tests comparing the number of BCTs between apps with higher and lower usability

Usability	n	Mean (SD)	95% CI		p-value
Higher	50	6.98 (2.82)			
Lower	15	7.20 (3.41)			
Difference	-	0.22	-1.52	1.96	0.802

## 8. What is the association between the app size and the number of BCTs?

The analysis was conducted separately for apps in iTunes and GP because the mean sizes across stores were different. T-tests were used to compare the mean number of BCTs for apps above and below the median size (82.2 MB in iTunes and 29.6 MB in GP).

Neither difference reached significance, but there were numerically more BCTs, on average, in the larger apps (Tables 21 and 33).

**a) The association between app size in iTunes and the number of BCTs**

Table 32: The results of the t-tests comparing the number of BCTs and apps above and below the median size (iTunes)

Size in iTunes	n	Mean (SD)	95% CI		p-value
<b>&gt;= 82.2</b>	23	7.87 (2.26)			
<b>&lt; 82.2</b>	22	6.5 (3.17)			
<b>Difference</b>	-	-1.370	-3.02	0.28	0.102

**b) The association between app size in GP and the number of BCTs**

Table 33: The results of the t-tests comparing the number of BCTs and apps above and below the median size (GP)

Size in GP	n	Mean (SD)	95% CI		p-value
<b>&gt;= 29.6</b>	19	8.16 (2.32)			
<b>&lt; 29.6</b>	18	6.78 (3.59)			
<b>Difference</b>	-	-1.380	-3.39	0.63	0.171

**9. What is the association between the number of features and the number of BCTs?**

A simple linear regression was used to estimate the relationship between the number of app features based on the number of BCTs. A significant association was found:  $F(1, 63) = 47.62$ ,  $p < .0001$ , with an  $R^2$  of 0.43. Each increase in the number of BCTs is associated with a 0.60 increase in the total number of features (SE = 0.09,  $p < 0.001$ ).

**10. Is there an association between the presence of peer-reviewed studies and the number of BCTs?**

For 12/65 apps there was evidence for presence of peer-reviewed studies related to the apps in the sample. For apps with evidence of peer-reviewed studies, the mean number of BCTs was 8.08. For those with no evidence of peer-reviewed studies, the mean BCT number was 6.79. An independent t-test showed there was a difference in the mean number of BCTs between the apps that had a peer-reviewed study compared to those that did not (-1.29, 95% CI: -2.51 to -0.08,  $p= 0.038$ ), with a significantly larger mean number of BCTs in apps with peer-reviewed studies (Table 34).

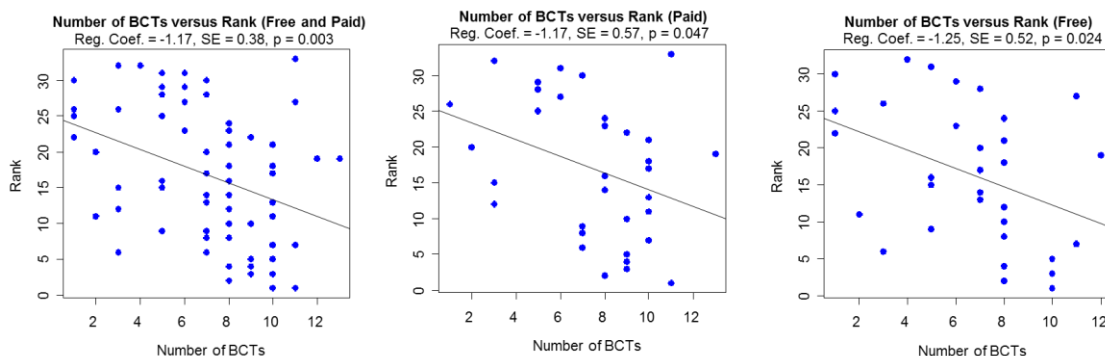
Table 34: The results of the t-tests comparing the number of BCTs between apps with and without evidence of the presence of peer-reviewed studies

Presence of peer review	n	Mean (SD)	95% CI		p-value
Yes	12	8.08 (1.44)			
No	53	6.79 (3.15)			
Difference	-	-1.29	-2.51	-0.08	0.038

**11. Is there an association between the app ranking and the number of BCTs?**

There was a significant linear relationship between the number of BCTs and store ranking. Overall, each additional BCT was associated with a mean decrease of -1.17 (lower ranks are better) in the ranking order (SE= 0.38,  $p = 0.003$ ). The effect was similar for free and paid apps (Figure 34).

Figure 34: The results of linear regression between the store rank and the number of BCTs



### 3.4.8 Sensitivity analysis

Sensitivity analysis was performed including only those BCTs that were classified as “present beyond all reasonable doubt” (++). The results yielded a consistent, slightly stronger result compared to the main analysis. There was no effect of the number of BCTs “present beyond all reasonable doubt” on user ratings for GP (OR= 0.99, CI 95% 0.90 to 1.08) but there was a significant effect for iTunes (OR= 1.18, CI 95% 1.08 to 1.29).

The sensitivity analyses conducted on the association between the self-regulation BCTs and user ratings including only those BCTs that were classified as “present beyond all reasonable doubt” [217] showed that for iTunes, the BCT 2.2 Feedback on behaviour crossed into significance. Otherwise, the results were consistent with the main analysis findings. The results of the sensitivity analysis are presented in Table 35.

Table 35: Sensitivity analysis including BCTs with only high level of confidence in the presence in the apps (++)

Variable	App store	OR (95% CI)	App store	OR (95% CI)
Group 1: Goals and planning	iTunes	1.16 (0.87-1.55)	GP	1.11 (0.83-1.49)
1.1. Goal setting (behaviour)	iTunes	0.91 (0.47-1.76)	GP	1.65 (0.85-3.20)
1.3. Goal setting (outcome)	iTunes	1.18 (0.64-2.18)	GP	0.81 (0.44-1.49)
1.4. Action planning	iTunes	1.07 (0.63-1.84)	GP	1.42 (0.84-2.40)
Group 2: Feedback and monitoring	iTunes	1.62 (1.22-2.13)*	GP	1.25 (0.95-1.64)
2.2. Feedback on behaviour	iTunes	2.07 (1.02-4.22)*	GP	1.76 (0.84-3.68)
2.3. Self-monitoring of behaviour	iTunes	1.57 (0.95-2.61)	GP	2.05 (1.23-3.39)*
2.4. Self-monitoring of outcome(s) of behaviour	iTunes	1.61 (0.93-2.77)	GP	0.98 (0.57-1.69)

Note: \*,  $p < 0.05$

#### 3.4.8.1 Analysis of apps that are appropriate for beginners

When restricted to apps appropriate for beginners, in iTunes each additional BCT increased the likelihood of receiving 4 or 5 stars by 16% (OR= 1.16, 95% 1.05 to 1.29), for GP the effect was not significant (OR 0.97, 95% 0.88 to 1.08). This result was consistent with the main finding.

### 3.4.8.2 Changing the threshold for outcome (star ratings – 5 starts vs 1, 2, 3, 4)

The association between the number of BCTs and user ratings using threshold of 5 versus 1,2,3,4 starts was used, the results were similar to the primary analysis. The model with the app store interaction did not converge hence each store was modelled separately (Table 36).

Table 36: Sensitivity analysis assessing the association between BCT and star ratings (5 versus 1,2,3,4 stars)

<b>5 vs 1,2,3,4 in iTunes</b>	<b>OR</b>	<b>95% CI</b>		<b>p-value for the interaction</b>
Higher rating	1.46	0.75	2.86	N/A
Lower rating	1.04	0.93	1.16	

<b>5 vs 1,2,3,4 in GP</b>	<b>OR</b>	<b>95% CI</b>		<b>p-value for the interaction</b>
Higher rating	1.48	0.96	2.28	N/A
Lower rating	0.98	0.92	2.28	

### 3.4.8.3 Excluding apps with less than 25 user ratings

For this analysis, 6 apps with <25 user ratings were excluded. For iTunes, the OR for number of BCTs was 1.08, CI 95% 1.00 to 1.17; for GP, the OR was 0.91, CI 95% 0.84 to 0.99. In this analysis, additional BCTs in iTunes increased the odds of a high rating, while in GP, apps with additional BCTs tended to receive lower ratings.

### 3.4.9 Retrospective power analysis

Based on simulations, this study had 97% power to detect an OR  $\geq 1.2$  for the effect of additional BCTs on high ratings with 64 apps. Since there was high power in this study and there was no significant result found in the overall models (number of BCTs and user

ratings), this suggests that the true increase in the chance of a star rating of 4 or 5 with each additional BCT has an OR of < 1.2.

## **3.5 Discussion**

### **3.5.1 *Principal findings***

This study explored which app quality indicators and app characteristics were associated with app popularity (assessed through user ratings) and likely efficacy (assessed through the inclusion of BCTs). The findings indicated that: 1) overall, there was no association between popularity and likely efficacy across both stores; 2) App size, the number of app features, the presence of peer-reviewed studies associated with the app, and the ranking in the apps stores was positively associated with the number of BCTs; 3) Expert involvement was associated with higher user ratings in iTunes, but not in GP. Apps that had privacy policy were more likely to receive higher user ratings in GP, but not iTunes. There were no other associations between the app quality indicators and characteristics and user ratings.

Some unexpected findings were that the user ratings differed by the app store. Specifically, 1) users in GP were more likely to award higher user rating than those in iTunes, 2) the effect of the number of BCTs was different across the app stores. Specifically, there was an effect of the number of BCTs for iTunes with each additional BCT increasing the likelihood of higher user rating in iTunes by 15%, however there was no effect for GP app store. Next, I discuss the findings of the study and follow with the consideration of the unexpected findings.

## **Finding 1: No association between popularity and likely efficacy of PA apps across both stores**

This study supports the findings of previous research which showed that apps that were highly rated, highly ranked or frequently downloaded were not of high quality in weight management [136, 172], smoking cessation [206] and alcohol reduction [138] apps. In addition, the inclusion of BCTs that have been shown to be effective in increasing PA, i.e., self-regulation strategies, were also not associated with higher ratings and this results supports the similar findings of Bardus, van Beurden [136]. These accumulating evidence suggests that popular apps are not necessarily of high quality. The consideration of the reasons for this findings is discussed below.

### **Why would almost 3 million users not be able to filter apps that are more likely to be effective?**

The summary of the distinction between the environments of the app market versus public health discussed in this section is presented in Table 37. I considered two main approaches to well-being as the basis to aid in understanding the differences between *liked* and *effective*. There are two main perspectives on well-being, the hedonic and eudaimonic approaches. The hedonic perspective focuses on happiness and states that the ultimate goal of one's life is to maximize pleasure and minimize pain. Hedonic enjoyment occurs whenever positive affect is accompanied by need satisfaction. Hedonism, as related to well-being research, involves both bodily and intellectual pleasures [218]. As 70% of adults in the UK report they would like to increase their PA levels [55] the app market may be a virtual place to look for their needs satisfaction. On the other hand, the motivation of the commercial app market is to produce a profit through app sales, in-app purchases, and any peripheral products associated with the app. In this context, the user is the consumer and the apps must be liked to be downloaded. I



argue that behaviour change may not be the primary motivator for why commercial PA apps are developed. In this context, the user's motivation to satisfy the immediate need might be short-lived and fail to produce a long-lasting behaviour change.

On the contrary, the eudaimonic perspective states that happiness does not equal well-being. Eudaimonia is characterised by engaging in purposeful activities that promote personal growth and facilitates realization of one's potential [219]. Eudaimonism [220] is a theory where the vital point of human existence is gaining the knowledge of daimon (true self) and living in consistency with one's values. This may mean abstaining from immediate need gratification if it is not consistent with self-identity. In the context of this study, what is effective might not necessarily be liked and vice-versa. The context is shifted from commercial market to public health and the user is not seen as a consumer per se but a citizen, a user of the healthcare system, a patient. Within this perspective immediate happiness may be substituted for deliberate engagement in health behaviours that, at first, may not bring immediate pleasure. This perspective produces long-term well-being. In this case, the decision to engage in health behaviour, such as PA, might be considered as not immediately increasing happiness but nevertheless consistent with one's values and identity.

Hence, targeting hedonic motivation (the immediate pleasant experience following engagement in PA) as well as the long-term influences on behaviour related to self-identity (eudaimonic motivation) would increase the behaviour change potential of apps for PA long-term.

Table 37: Differences between liked / popular and effective apps

Differences between	Liked / popular	Versus	Effective
Setting	Commercial app market		Public health
End-user	Consumer		Citizen/Patient
Motivation	Hedonic perspective		Eudaimonic perspective

**Finding 2: App size, the number of app features, the presence of peer-reviewed studies associated with the app, and the ranking in the apps stores positively associated with the number of BCTs**

Across behaviour change interventions, the relationship between app quality and popularity is mixed. Whilst Azar, Lesser [172] found that apps with high **rankings** were of low quality, Chen, Cade [137] reported a positive association between ranking and the number of BCTs, both in a sample of weight management apps. A positive association between app ranking and the number of BCTs was also found in this study. It is unknown what constitutes the app ranking algorithm but variables that have been indicated to contribute are the number of downloads and user ratings [198]. As app features were associated with the number of BCTs, as found in Bardus, van Beurden [136], it is possible that apps which provide users with more options may be more likely to be downloaded, if not used.

However, in this study, more BCTs were associated with both more app features and a larger app **size**. This finding might suggest that apps larger in size, although providing more BCTs, might be deleted earlier from the device, as high memory allocation is the third main reason why users uninstall their apps [221]. Previous studies found that, although apps with higher user ratings had higher usability [147], more features in apps decreased their usability [203].

Hence, the inclusion of BCTs or features should be limited to those that are necessary. This could be achieved by personalising the app content to users' needs. For this reason, systematic development of the apps based on user needs, grounded in behaviour change theory, with systematic selection of those BCTs that are most likely to be effective is important. In addition, as any behaviour change process is non-linear, dynamic and responsive designing tailored interventions adapted the user needs might be more effective [222] and engaging [223] than providing a one-size-fits-all intervention for increasing PA. For example, goal setting theory states that the goals should be evaluated periodically against the performance and adjusted accordingly if there is such need [224, 225].

In this study, apps that had peer-reviewed publications associated with the app included more BCTs in comparison to those that did not have any studies associated with the app. This finding should be interpreted with caution as only 12 apps had a published article in the scientific database. However, this might suggest more prominent inclusion of app features that address behaviour change in those apps for which a peer-reviewed study exist.

This study found that there was no relationship between ranking and star ratings. Although the variables used for the ranking algorithm are not known, user feedback is considered, and expected, to have an impact on the app position in the ranks [198]. The potential reasons for this finding might be that, as this study aimed to look at popular apps, only "top-ranked" apps were selected for the study, and there might have not been enough variability in the sample to find an association because there were no unpopular and low-ranked apps to form a contrast. In addition, dichotomisation of the star ratings might have been too crude an outcome measure to assess how feedback plays into

ranking considering that only 12% of ratings were 1, 2, or 3. An alternative reason might be that the user voice in the form of user ratings is down-weighted in the app rankings produced by the app stores. As iTunes and GP are both commercial environments, the possibility of app companies purchasing their ranking position cannot be excluded.

**Finding 3: Expert involvement was associated with higher user ratings in iTunes, but not in GP.**

When considering the predictors of higher user rating in the app store, only one association was found. Those apps for which there was evidence of expert involvement were more likely to receive higher user rating in iTunes, but not in GP. These results have to be interpreted with caution as only 12 apps had evidence for the involvement of experts. Similarly, Pereira-Azevedo, Osório [193] found that download rates were higher when experts were involved in the development of urology apps. No other quality indicators and app characteristics predicted high user ratings.

**3.5.1.1 Unexpected findings**

**Finding 1: users in GP were more likely to award higher user rating than those in iTunes**

**Finding 2: the effect of the number of BCTs was different across the app stores**

The findings relating to the difference between the stores and user ratings were not previously reported. There are various possible explanations for this difference, including demographic differences in the population of users of each store [226, 227]; differences in the review and approval processes between the two stores [221], or the play of chance. Next, I will discuss these in detail.

Market research reveals some demographic differences between iPhone and Android users. There is a substantial difference in household income between iPhone and Android users, with Android dominating towards the bottom of the income ladder, and iPhone usage increasing with income. iPhone users have, on average, higher level of education than Android users. In addition, the research found that 40% of Android users reported having an immediate family member with heart disease, in comparison to 25% for iPhone users [228]. iPhone users are more likely to make purchases using e-commerce and they are likely to make more expensive purchases than Android users [227]. In 2015, users of iTunes spent \$20 billion dollars and 4 times more per user than GP users [229]. iOS users engage with their phones more than Android users and are likely to also own another Apple products which may suggest their loyalty to the brand [230]. In addition, a recent study found evidence for prevailing stereotypes between users of the iPhone and Android. In a survey of 240 people, those that use Android were perceived as more honest, humble, agreeable and open and seen as less extroverted than iPhone users [231]. However, a recent study comparing the users of iOS and Android on key personality traits found small to negligible differences [226].

In addition, as iPhone tend to be more expensive than Android, it is possible that iPhone users have higher expectations, reflected in their tendency to assign lower ratings, in comparison to Android users. Moreover, this study showed that apps in iTunes are substantially bigger than apps for Android (82.2 MB versus 29.6 MB, respectively) and this might influence the lower ratings of the apps in iTunes if iPhone users are able to download a smaller number of apps than Android users due to size restrictions.

Lastly, profit as a motivation and the app development and app review processes might influence the observed difference between the app stores in this study. The market share of Android is domineering the mobile landscape (85.9%), with iOS having a 14.1% share.

However, when considering profits, Apple dominates. There are many software and device variants for Android [232]. Contrary, iOS app development tends to be more efficient due to homogeneity of the iOS platform. In addition, Apple has a stricter review process than GP, both for developers attempting to get their product onto the Apple app store, and users who want to rate the apps. This might explain why there were 25 times more user ratings given by Android users compared to iPhone users. Profitability, relative ease in app development, and the Apple review process are factors that could explain the higher quality of apps being produced for iOS [233]. This may also explain why the effects of the content of the apps (number of BCTs) has a different relationship to user ratings across the stores.

Last, it is possible that the findings occurred by chance. Although the test for the interaction was significant, the CIs seem to overlap 1.15 for lower (1.06) and upper bound (1.25), hence there is a possibility that this result may not represent a true difference. It is unlikely that for iTunes users there is an effect of BCTs on user experience whilst there is none for GP users.

### **3.5.2 Strengths**

This is the first study that aimed to explore what predicts higher user ratings in popular PA apps using crowdsourced data from the apps stores. The main strength of the study includes the use of systematically derived quality indicators based on assessment conducted by two reviewers. The sample was identified from the most popular publicly available apps from 2 major app distribution platforms. The use of BCT Taxonomy provides a standardised assessment tool which has been utilised in other studies assessing the content of apps [137, 139, 145, 176]. In addition, the primary indicator of the likely effectiveness of the apps included 2 operationalisations (the number of BCTs and the specific BCTs that have been shown to be effective in increasing PA) which increases the robustness of the results. Last, Guzman [234] argued that the star rating

represents an average score for the whole app that combines both positive and negative evaluations aggregated across users. The study, however, used rater-level data which included individual ratings from 2.8 million users. These large numbers mitigate the problems posed by averaging the star rating across users.

### **3.5.3 Limitations**

This study has some important limitations. First, the main limitation of the study relates to the variables used. It is possible that user ratings, as expressed by the stars assigned to the apps, can relate to different aspects of the app functioning and content. There is evidence suggesting that app reviews tend to occur near a new release, which may suggest that the ratings reflect comments on the specific updates of the software [235]. In addition, the possibility that user ratings were influenced by fake reviews cannot be excluded [148, 149]. Moreover, some developers used the *power of defaults* by asking the user to “give the app 5 stars”. However, the user rating was considered the most appropriate measure to use since it represents a user-led feedback that reflects user experience. Similarly, the choice of the BCTs as an approximation for likely efficacy was selected because studies assessing the efficacy of the apps on the market are scarce [84, 236, 237]. It is likely that the quality of BCTs implementation will have an important role for effectiveness. Second, the ranking algorithm from which the sample was derived is unknown. Hence, this lack of transparency prevented evaluation of how the calculation of rank might have influenced the app selection. However, apps appear in rank order by default in app stores, and so the rank affects what users are seeing. As this study was assessing the most popular apps, the choice of highly ranked apps was considered the most appropriate for the context of this research. Third, the GP market tend to have more ratings than iTunes. This is most likely because there is more GP users. In addition, the process of app review is more complex in the latter [238]. This was addressed in the study by using the weighted averages of the ratings across the stores and by controlling for store in the regression models. While there is a difference in the app review process

in both app stores, feedback from both stores should be recognised as valid and important. Fourth, as there is very little known about what predicts what users like in apps, this study was exploratory in nature, and the p values are descriptive and aimed for hypothesis generating rather than generation of the findings to other apps. As such, there was no a priori type 1 error control for multiple comparisons so p values should be treated as descriptive and inference should be made with caution. Fifth, the difference in star rating by store was also unexpected, and hence the app store interactions with covariates of interest were a post-hoc addition to this plan. Sixth, there were only 12 apps that had evidence of health expert involvement, hence the significant association between expert involvement and user ratings in iTunes should be interpreted with caution. Seventh, there is evidence that some BCTs may operate in clusters [239]. However, the analysis of the individual BCTs within the groupings was not adjusted for other BCTs within the cluster. Last, failure to detect the skew in the original primary outcome used to power the study is a limitation, however, a retrospective power analysis showed the study had power to detect an OR of 1.2. As no significant result were found, the true OR is likely to be  $< 1.2$ .

### **3.5.4 Implications**

This study showed no evidence of an association between popularity and likely efficacy. The findings suggest that, at present, allowing the commercial market to determine which PA apps are downloaded is unlikely to be an effective method of public health promotion in terms of increasing overall levels of PA.

#### **3.5.4.1 Implications for research**

As this study showed that BCTs may increase the size of the apps, and app storage is an important factor influencing the decision to remove an app, only the BCTs that are most likely to be effective and liked should be included in the apps. Further research is needed to understand which BCTs, and in which combinations, are most effective in



increasing PA when delivered via an app; how these BCTs can best be delivered; and how to combine features which promote effectiveness with those that promote popularity. Last, the differences in iTunes and GP users is an unexpected finding and not one that the study set out to identify, (i.e. not an *a priori* hypothesis). Future researchers should be aware of the potential differences between iTunes and GP users and ensure research is carried out on both platforms.

#### **3.5.4.2 Implications for policy**

This study showed no evidence of association between what is liked by users and likely efficacy. In addition, app popularity was not associated with other quality indicators assessed, such as the presence of peer-reviewed studies. Moreover, assessing whether non-commercial affiliation and user involvement influenced ratings was not possible as only one app had evidence for non-commercial affiliation and there was no evidence for user involvement in the app development.

As such, the findings have some implications for public health policy. Apps aimed to increase PA represent the largest category in the two major app stores [240] which illustrates public demand for engaging in PA. The findings of this study suggest that, at present, allowing the commercial market to determine which PA apps are downloaded is unlikely to be an effective method of public health promotion in terms of increasing overall levels of PA.

Initiatives to identify and promote high quality apps have been in development. The recently reopened NHS App Library provides a database of 72 apps that fulfilled a series of criteria consisting of: 1) digital assessment process (conducted by the developers or an independent organisation), 2) content review which involves a more detailed assessment of the health and behaviour change content used [241]. However, at the

time of writing, the database's section "healthy living" included only two apps focussing specifically on PA (both developed by PHE).

### **3.5.5 Conclusions**

To date, this is the first study to assess the association between popularity (measured using user ratings of the apps) and likely efficacy (measured using the inclusion of the BCTs) of publicly available highly-ranked PA apps available in the major app stores. No relationship was found between popularity and likely efficacy suggesting that popularity does not assure high quality, and what is liked may not be what is likely to be effective. Thus, promotion of public health is unlikely to be achieved by allowing market forces to determine which PA apps are used. Both researchers and developers should consider the potential differences between iTunes and GP users and ensure research is carried out on both platforms. More studies are needed to assess the effectiveness of apps with users in a real-world setting to investigate the app components that are both effective and valued by the users.

#### **Citation for the published peer-reviewed article for this study**

Bondaronek, P., Alkhalidi, G., Slee, A., Hamilton, F. L., & Murray, E. (2019).

Relationship between popularity and the likely efficacy: an observational study based on a random selection on top-ranked physical activity apps. *BMJ Open*, 9 (11), e027536. <https://doi.org/10.1136/bmjopen-2018-027536>

See Appendix R for the published peer-reviewed journal article.

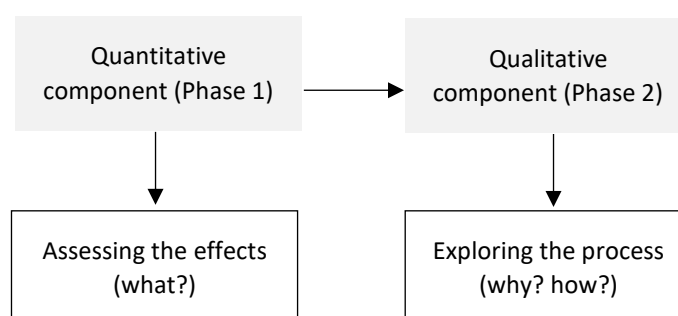
## **PART 2: FOCUS ON THE USERS**

### **The feasibility and acceptability of a randomised crossover trial assessing two popular physical activity apps: a mixed methods study**

In Part 1 of my PhD I assessed the quality of selected highly popular PA apps on the market and explored if user ratings could be used to predict what users value in those apps. In this part of this thesis I address the question of whether these apps have the potential to work, i.e., if they can lead to behaviour change.

Part 2 is a mixed-methods design consisting of a quantitative component assessing the feasibility of a future fully-powered RCT, acceptability of the trial methods and the potential effectiveness of two popular and highly ranked PA apps (Phase 1). However, a quantitative study should be accompanied with a more in-depth contextual exploration of how and why users choose to interact with the technology and what (if any) aspects and features of the app they considered as helpful/unhelpful. The qualitative component (Phase 2) explored the acceptability of the trial procedures and the experience of using the PA app interventions.

Figure 35: the overall structure of this sequential mixed-methods study



## **CHAPTER 4. Methodological considerations: the feasibility and acceptability of a randomised crossover trial assessing two popular physical activity apps**

### **4.1 Chapter overview**

This chapter reports on the methodological considerations when designing the quantitative component of this study which was a feasibility crossover trial of two popular apps on the market. The background included the consideration of what a high quality trial is based on the Risk of Bias assessment tool.

### **4.2 Background**

#### **RCT is the gold standard**

Randomised controlled trials are considered the gold standard [242] for assessing efficacy or effectiveness (in a pragmatic RCT) because randomization minimises the impact of confounding variables that may influence the results [243]. In this Background section to this study, I first present the design of this study, following by a consideration of the components of a high quality trial, and how I approached the process to design the best study achievable with the resources available to me.

\_\_\_\_\_ Reflective log \_\_\_\_\_

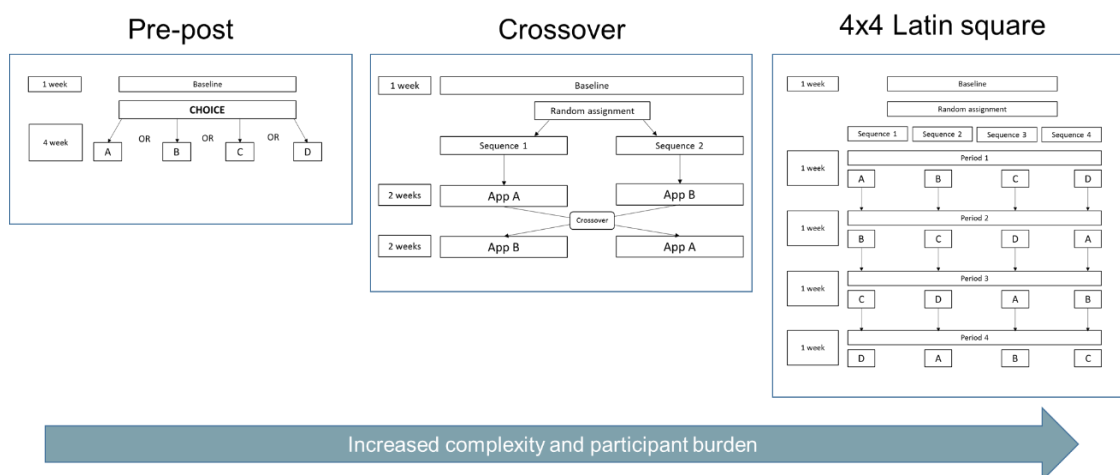
#### ***Making balanced compromises: selecting the design to assess the potential effectiveness of the apps:***

In this section I present the decision-making journey around the design of this study. As reported in Chapter 2, in my review (Study 1) I assessed 65 apps. In an ideal world, I

would consider conducting 65 RCTs, all with a control arm. This approach would be extremely resource intensive, of both time and funding. Still, I was determined to assess 4 apps as the results of Study 1 showed four types of apps that were the most common on the app market (the workout apps, tracking of movement, running programmes, and pedometer-based apps).

The factors I considered to be important for the design of this study were the balance between the complexity of the design and the increased burden for the potential participants. Figure 36 visualises the thinking process (in a simplified linear way) I followed to reach the decision to conduct a crossover trial. The table with the consideration of other designs are found in Appendix I.

Figure 36: The process of negotiating the design for this study



First, I considered a non-randomised pre-post study examining within-participant changes in PA. The advantages of this simple design include the ability to closely mimic the app store as I would be able to “offer” participants a choice of four apps and then follow them to assess if their PA improved. In addition, this design was both simple and

least burdensome for the participants. However, I quickly realised that I was addressing a different research question to the one I aimed to investigate. The research question would be not that of potential effectiveness of the apps. It would help answer a question: “When given a choice of four apps, which apps do participants choose?” In addition, without a reference to a comparison group, this study would be difficult to interpret and may lead to erroneous conclusions about the effectiveness of the interventions.

I was excited to learn about the Latin square design. Using this design, I could assess four apps with all participants by randomising participants to one of the four sequences of app assessment. However, I realised, with the pain of letting go, that a trial of this kind would be very complex where I would “force” participants to use all four apps. The trial fatigue was expected to be high and I could expect strong period effects where the sequence of app assessment would influence the completeness of the data.

I had to use some strong negotiation with myself to reach a consensus that was somewhat in the middle – the crossover design. This was a way to simplify the 4x4 Latin square. I had to consider two apps rather than four, but I thought this design would be possible to execute and would decrease the participant burden and the potential trial fatigue. Simultaneously, I would still be able to randomise participants to the sequence of two apps hence increasing the validity of the study. In addition, by using a crossover design, each participant would act as their own control, hence I would be able to assess the overall effect of the apps but also to compare the two apps. This was certainly not an ideal method but it was an optimal way when considered the resources I had at my disposal: intellectual, time, and finance.

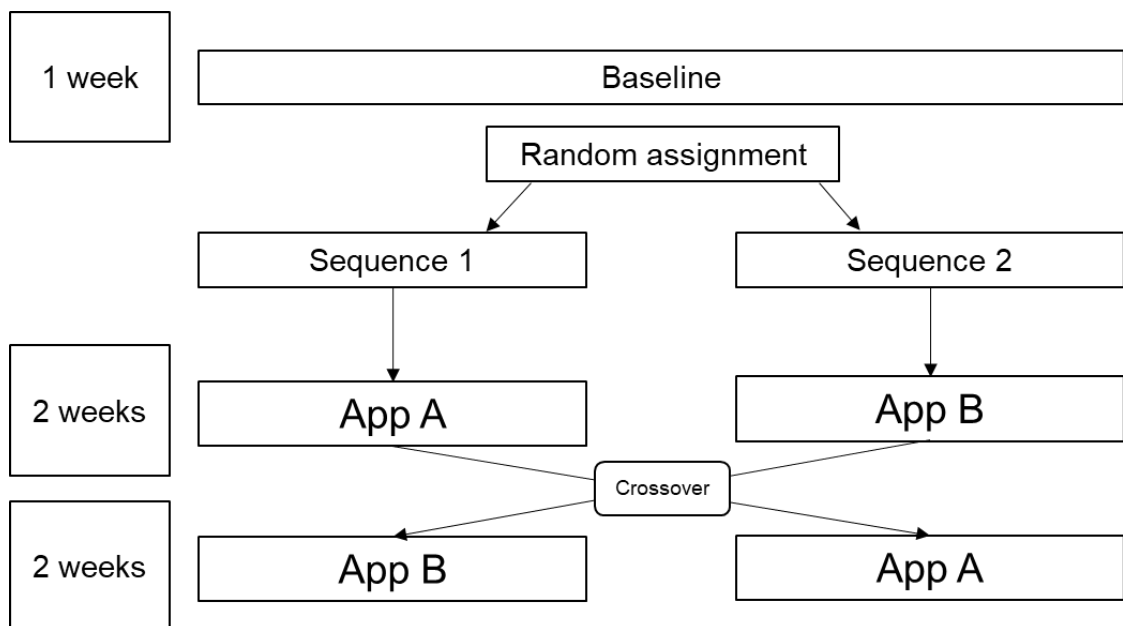
---

End of the reflective log

---

Figure 37 presents the design of this feasibility crossover trial. The trial lasted 5 weeks (1 week baseline, 2 weeks intervention period for each app). First, participants' baseline PA was assessed as, in crossover designs, participants' baseline PA can be compared to their follow-up outcomes (participants act as their own control arm).

Figure 37: Randomised crossover trial design used in this study



### What makes a high quality RCT?

In the next section, I discuss the main components of a high-quality trial including the importance of a feasibility study, and the methods to increase the internal and external validity of the trial including the rationale for the decision I made when planning this crossover trial.

The MRC framework for developing and evaluating complex interventions argues that before a definitive fully powered trial is conducted, it is crucial to conduct assessment of the feasibility and acceptability of the study [107]. Feasibility determines if the study

design, procedures, and the intervention can be executed by the researcher. Acceptability assesses the suitability of the study design, procedures, and the intervention from the perspective of the participants and intervention deliverers [244]. Hence, this study was a mixed-methods feasibility study to inform a decision about whether to proceed to a large-scale definitive randomised controlled trial (RCT).

### **Internal validity**

To design this study, the Cochrane Risk of Bias (RoB) tool for randomised trials [245] was used to determine which factors impact the quality of RCTs, and hence, to increase the internal validity of this study. Each of the biases described in the quality tool is considered in the next section.

**Bias arising from the randomisation process** relates to the randomisation of the participants, the concealment of the allocation sequence, and the occurrence of imbalances in the baseline characteristics that may influence the results.

In this study, April Slee (a PhD student in the department), provided me with her statistical advice. Specifically, she helped me to devise the statistical analysis plan based on the objectives I developed; provided guidance on running the analyses on the statistical software when needed; and developed and managed the random sequence generation.

In addition, the differences between baseline characteristics of the participants were assessed. Randomisation eliminated systematic differences in these variables, but cannot avoid randomly occurring differences. Socio-demographics and other baseline characteristics, such as intrinsic/extrinsic motivation for PA [246, 247] can have an impact on the outcomes measured [248]. To minimise the impact of these variables that



were not targeted in the intervention, they were assessed at baseline and the differences between the randomised groups were assessed to ensure that the imbalances between the baseline characteristics were accounted for (if such were found).

**Bias due to deviations from intended interventions** relates to the blinding of participants and researchers to the intervention group randomisation. It is difficult, if not impossible, to blind the participants to the allocation in behavioural interventions. Although I did not provide the names of the apps used in the intervention in advance, so as not to influence expectations, the concealment of the intervention was not possible.

The most recent RoB (2.0) acknowledged that the tool was limited to studies where blinding was possible. The new RoB tool expanded this section including the potential confounding factors, one of which was relevant for this trial: non-adherence to intervention. It is possible that the participants' pre-conceived attitude towards an app could influence their adherence to the intervention. The main analyses were performed using the Intention-to-Treat (ITT) method which can estimate the effect of the assignment to the intervention [245].

**Bias due to missing outcome data** relates to the potential imbalances in the outcome data due to attrition. This assessment includes the consideration of how much data are missing and why. As this is a feasibility trial, the investigation of this particular bias formed the primary outcome, which was assessed as the retention rates and the completeness of the data.

To reduce missing data in this study, **incentives to increase completeness of the data** were used. An expert consensus workshop findings [249] and qualitative study[250] suggested that the monetary incentives for participation in an RCT should be neither too

high as to be perceived as coercive, nor too low so that participants feel undervalued. The range of the monetary incentives suggested was £5 to £20. As this study lasted 5 weeks with potential qualitative interview I decided that £20 was an appropriate token of appreciation for participants' time.

In addition, as this study aimed to explore the short-term effect of the apps a two-week assessment period was selected to maximise data completeness and decrease participant burden.

**Bias in measurement of the outcome** relates to the potential confounding effect of the measurement used, whether influenced by the assessors or the participants themselves. To minimise this bias, the primary outcome for the potential future definitive trial was measured objectively using an accelerometer device.

#### **Objectively measured PA - accelerometer**

There is no standard method to measure PA [251, 252]. However, an accelerometer (a motion sensor) can be used to assess the duration and intensity of activity, and it is considered to provide the most valid and reliable measurement of PA [253]. Previous studies assessing the efficacy of PA interventions (both using digital intervention and face-to-face methods) using accelerometers tended to have high standard deviations and many of the studies were underpowered [187, 254-256]. An example was Wang, Cadmus-Bertram [256] where the authors expected to see 17% increase in total PA levels but saw only a 0.3% increase.

## **IPAQ**

As this was a feasibility trial, I explored the optimal method for assessing the outcomes in the definitive trial, hence a subjective measure was also used to assess PA using the International Physical Activity Questionnaire (IPAQ) Short Form [261].

## **Additional outcomes**

As the intervention period was short, I considered that it was important to assess the psychological predictors of PA. The basis for the selection of the psychological predictors of PA to be measured in the study was a recent Lancet review [263] which synthesized 9 systematic reviews of the determinants and correlates of PA in adults. I followed the steps described below. First, I summarised the determinants/correlates in a table (Appendix J). Second, if applicable, I mapped the determinants/correlates onto the COM-B and TDF frameworks with a rationale to either measure these variables in the quantitative study or to explore these in the qualitative study (Appendix K). Third, for each of the determinants/correlates, I considered whether they could be modified in the intervention within the 2-week intervention period.

Based on this process, the PA determinants/correlates that I selected to be measured in the quantitative study were self-efficacy and intentions. When I investigated other psychological determinants of PA, I found that different studies assess various psychological constructs related to PA, such as intrinsic/extrinsic motivation, optimism, positive and negative affect. However, a frequently measured variable I found to be used is the expectations of the outcomes of conducting PA [264, 265]. The measure of expected outcomes, defined as the belief in the likelihood that a specific behaviour will be followed by a particular outcome [266], is included in various social and cognitive models, such as Self-efficacy theory [267], Transtheoretical Model [268], and Theory of Planned Behaviour [102]. However, this variable was not included in the review of

reviews which I used to systematically select the measures for this study. This might be because of the different conceptualisation of the construct in various studies [269]. I decided it was important to include a measure of the PA outcome expectation in the study.

**Bias in selection of the reported results relates** to the selective reporting of the outcomes and using multiple analyses of the outcome data. In this study, a protocol was followed. However, as the purpose of this feasibility study was to explore the most optimal method of assessing two apps for PA, post-hoc analyses were conducted. An example of this includes the exploration of the variability of the PA outcome as presented in the results, in section ‘Variability in the outcome measure’, in Chapter 5.

In addition, issues related specifically to crossover trials were considered as described by the RoB 2.0 working group [270].

**Carry-over relates to the effect** of an intervention to produce a long-lasting modification, and hence crossover designs may not be appropriate in the presence of carry-over effects as the sequence of interventions is likely to matter. Considering the nature of a behaviour change interventions in this study, the carry-over effect was considered.

The intervention period lasted two weeks as the question of the study was not to assess the formation of a habit using the apps but rather the potential for effectiveness. Empirical study assessing habit formation in different health behaviours estimated that it takes 18 – 254 days to form a healthy habit, with simple singular behaviours, such as drinking more water, taking less time [271].

Hence, it was considered that a two-week period would not produce a persistent behaviour change.

Lastly, the rationale for the use of the crossover design was to mimic the behaviour of the user in the app store who might download multiple apps in a sequence to attempt to increase their PA as the statistics on app usage show that up to 90% of apps that are downloaded apps are used only once [147]. Most importantly, the question of whether the sequence of app assessment influences the outcome variables is a testable hypothesis which was examined in the study by assessing the period effects analysis, as recommended by the RoB 2.0 working group [270].

### **Consideration for ecological validity**

As this study aimed to explore the public health potential of PA apps, the considerations for external validity were emphasized. These included the focus on effectiveness rather than efficacy and population targeted.

This study assessed the potential for effectiveness of the apps. Hence, participants were not provided with instruction of the frequency of the app use. They were asked to use the apps in a self-directed way. This was to increase the applicability of the finding to the “real world”. However, a sensitivity analysis was conducted including those participants that engaged with the app intervention. As there is no agreement on what constitutes minimum engagement with a digital intervention, an arbitrary measure was used which was derived from the data. Participants’ *amount of use* variable of the app in minutes (on the digital behaviour change intervention Engagement Scale[272]) were split by median and those with median or higher amount of use were defined as those that engaged with the intervention.

The population targeted in this study comprised people classed as engaging with no or low levels of PA as, from the public health approach, the priority is to increase PA in those who are inactive [273]. As described in the inclusion criteria, having a low level of PA was the main inclusion criterion in this study. In addition, the strategy for recruitment was to obtain a diverse sample in order to increase generalizability of the results. People with lower digital health literacy were less likely to take part in the study [274]. To improve access to the study for people with lower digital literacy levels, organisations that aim to increase digital inclusions (e.g., Doteveryone) were contacted and asked to advertise the study.

Moreover, weather and seasonality impact level of PA [275, 276]. As the study commenced recruitment in January, sensitivity analysis was conducted to investigate if weather influenced the daily PA levels. In addition, as App B aimed to increase running which most participants would likely perform outdoors, the differences in PA between apps by weather were explored.

The UK Met Office provided the data for the purpose of this analysis. Hampton Water Works, Heathrow, High Beach, Kenley Airfield and Northolt are the closest weather station in London. As only Northolt recorded the snowfall values at the time of the trial running, the MET Office information specialist advised to use that dataset from the Northolt weather station. Precipitation was classed as rain (any daily rain) versus no rain and snow (any daily snow) versus no snow.

#### **4.2.1 Study aims and objectives**

The aim of this study was to investigate the feasibility and acceptability of a study assessing two selected PA apps to inform the design of a definitive RCT, and to assess

the effects of the app interventions on PA. The specific objectives are presented in table 38.

Table 38: Primary and secondary objectives of Study 3 and 4

Primary objectives	To determine the feasibility and acceptability of the app interventions and the trial procedures
	<ol style="list-style-type: none"> <li>1. To assess recruitment and retention rates in the trial</li> <li>2. To determine the acceptability of the trial procedures and the app interventions</li> </ol>
Secondary objectives	To explore the potential effects of the 2 selected PA apps on PA and psychological outcomes
	<ol style="list-style-type: none"> <li>1. To assess the potential effects of 2 selected PA apps on PA</li> <li>2. To assess the difference between the effects of each app on PA levels</li> <li>3. To assess the change in exercise self-efficacy, intentions, motivation, PA outcome expectancy, and mood</li> <li>4. To assess usability and engagement with the apps</li> <li>5. To inform sample size calculation for definitive trial</li> </ol>

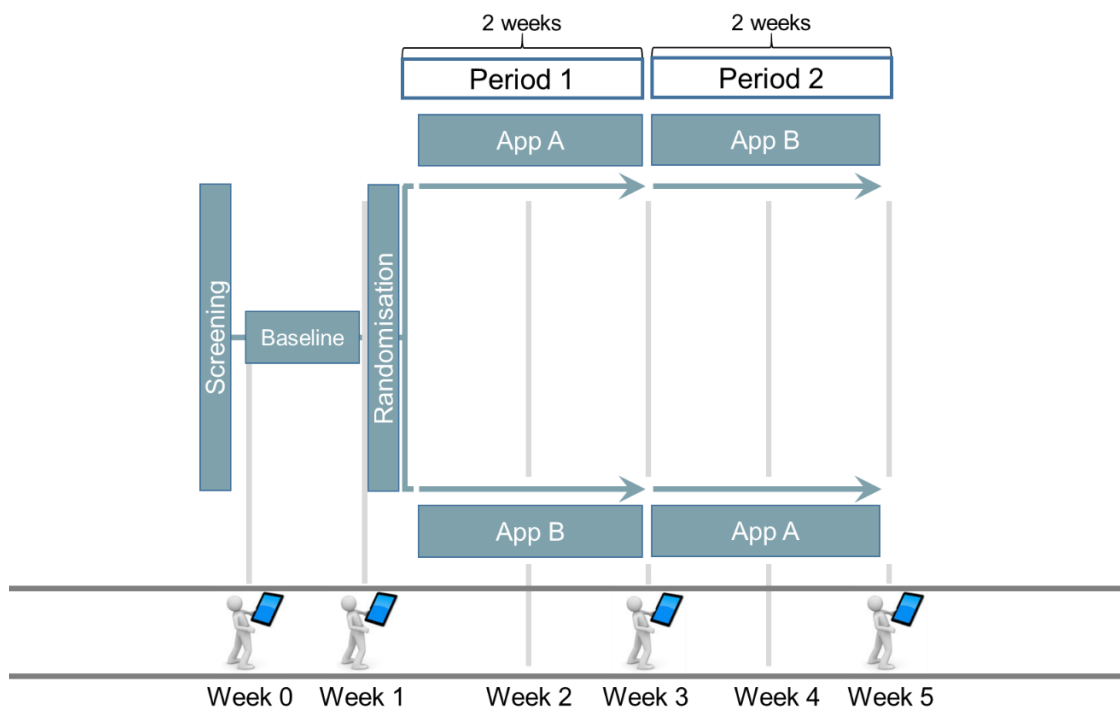
## 4.3 Methods

### 4.3.1 Design

The Consolidated Standards of Reporting Trials (CONSORT) guidelines for reporting pilot and feasibility trials and eHealth trials were used to report this study protocol [277, 278].

This study was a sequential mixed-methods feasibility study consisting of a randomised crossover trial (phase 1, Study 3), assessing 2 selected PA apps, and subsequent semi-structured interviews (phase 2, Study 4). The crossover trial allowed me to determine the practicality of the proposed study design, procedures and interventions, enabled me to explore the potential effects of the apps on PA, and informed the potential future definitive RCT. The study design schema (phase 1) is presented below (Figure 38).

Figure 38: Study design schema (phase 1)



#### 4.3.2 Setting and population of interest

Adults aged 18+ years who resided in London (UK) and surrounding areas were recruited for this study.



## **Eligibility criteria**

### *Inclusion Criteria*

- adults ( $\geq 18$  years' old) identified as “moderately inactive” or “inactive” using General Practice Physical Activity Questionnaire (GPPAQ)
- as the use of accelerometer requires some maintenance from the researcher, it was necessary to include only users that reside in/around London for practical reasons
- those owning a “smartphone”, iPhone (operating iOS 6.0 or newer) or Android (version 2.3.3 and up)

### *Exclusion criteria*

- do not speak English
- previous use of the apps of interest
- medical conditions that require special attention when conducting PA
- current participation in another research study that targets behaviour change
- unwilling to use the accelerometer as per study instructions The rationale for this exclusion criterion is that as studies assessing digital interventions are likely to have high attrition rates [279, 280], only participants who were more likely to adhere to the study instructions were included.
- unable to perform basic functions relating to app usage: download and/or navigate the app
- one week after the study recruitment commenced, it appeared that One You Couch to 5K was disabled on Motorola phones and this was confirmed by the developers at PHE. Hence, participants that used Motorola phone model were excluded.

### **4.3.3 App selection criteria**

The selection of the apps for the assessment in the study was based on the review and content analysis of the most popular apps in the market where 65 apps were assessed (Study 1). The criteria for the app selection for the trial was based on the trial population and followed these sequential steps:

- 1) The apps that might be appropriate for those who engage in no or low PA were included. Based on this criterion, 17 apps were excluded. The examples of the apps that were excluded were running programmes that prepare for 10 km runs or longer runs; apps that target bodybuilding using various equipment; interval timer apps as these do not provide PA content but only enable the user to time their personal workouts.
- 2) There were five app types identified in the review and content (see section 'Type of PA apps', Chapter 2). Running programmes and workout apps were considered for evaluation in this study. The rationale was that these apps provide specific instructions on how to perform the PA and often include demonstrations. As such, they require no prior knowledge of PA which is most suitable for those that engage in low levels of PA.
- 3) Apps had to be available on both iTunes and GP to be considered for the trial. The workouts and running programmes were then sorted according to the number of BCTs in the apps. The app with the most BCTs from the workout apps was *7 Minute Workout Challenge* by *Fitness Guide Inc.*, and the running programme with the most BCTs was *One You Couch to 5k* by *PHE*. Both apps provide one type of PA session (as opposed to different PA options) and thus would be appropriate for the assessment of the same PA behaviour in all participants in this study. Hence, these two apps were selected to be assessed in this study.

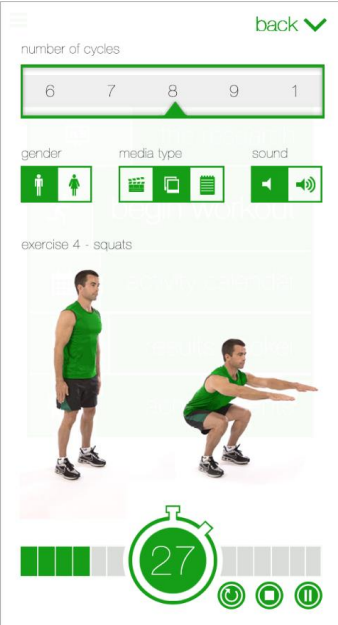
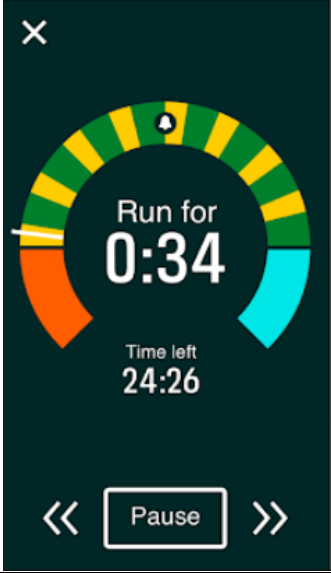
#### **4.3.4 Interventions**

Template for Intervention Description and Replication (TIDieR) [281] was used to describe the interventions: App A: Workout app (*7 Minute Workout Challenge 7 Minute Workout Challenge by Fitness Guide Inc.*) and App B: Running programme app (*One You Couch to 5k by PHE*), Table 39.

Table 39: TIDieR template for the study interventions

1. Brief name	<b>7 Minute Workout Challenge</b> by <i>Fitness Guide Inc.</i>	<b>One You Couch to 5k</b> by <i>PHE</i>																																				
2. Why	The app combines aerobic and resistance training into short, 7 minute HIIT targeting the whole body. Developers of the app cite the research of Klika and Jordan [185] as the rationale for the 12 exercises included in the app intervention	The goal of the app intervention is to get a user who is inactive to run 30 min (equivalent to around 5 km) after 9 weeks of training. The goal is achieved by incremental increases in walking to running ratio																																				
3. What – material	The content of the app is described using the major features (described in section ‘App features and mapping of the features and the BCTs’, Chapter 2) and the BCT Taxonomy (v1) [168]																																					
	<p>Major features of the apps include: Reminders, Logs, Automatic tracking, PA education in a variety of formats, Gamification</p> <table border="1" data-bbox="517 778 1267 1329"> <thead> <tr> <th data-bbox="517 778 786 818">BCT name</th> <th data-bbox="786 778 882 818"></th> <th data-bbox="882 778 1267 818">Evidence</th> </tr> </thead> <tbody> <tr> <td data-bbox="517 818 786 890">1.1 Goal setting</td> <td data-bbox="786 818 882 890">++</td> <td data-bbox="882 818 1267 890">The exercises are defined and set for the user</td> </tr> <tr> <td data-bbox="517 890 786 962">2.2 Feedback on behaviour</td> <td data-bbox="786 890 882 962">++</td> <td data-bbox="882 890 1267 962">The calendar records the exercises done</td> </tr> <tr> <td data-bbox="517 962 786 1098">2.4 Self-monitoring of outcome(s) of behaviour</td> <td data-bbox="786 962 882 1098">++</td> <td data-bbox="882 962 1267 1098">The option to record and track the weight changes</td> </tr> <tr> <td data-bbox="517 1098 786 1217">2.7 Feedback on outcome(s) of behaviour</td> <td data-bbox="786 1098 882 1217">++</td> <td data-bbox="882 1098 1267 1217">The graph shows weight change (inputted by the user)</td> </tr> <tr> <td data-bbox="517 1217 786 1329">4.1 Instruction on how to perform the behaviour</td> <td data-bbox="786 1217 882 1329">++</td> <td data-bbox="882 1217 1267 1329">The exercises are explained in text and audio</td> </tr> </tbody> </table>	BCT name		Evidence	1.1 Goal setting	++	The exercises are defined and set for the user	2.2 Feedback on behaviour	++	The calendar records the exercises done	2.4 Self-monitoring of outcome(s) of behaviour	++	The option to record and track the weight changes	2.7 Feedback on outcome(s) of behaviour	++	The graph shows weight change (inputted by the user)	4.1 Instruction on how to perform the behaviour	++	The exercises are explained in text and audio	<p>Major features of the apps include: Reminders, Logs, Automatic tracking, PA education in a variety of formats</p> <table border="1" data-bbox="1290 746 2040 1321"> <thead> <tr> <th data-bbox="1290 746 1559 786">BCT name</th> <th data-bbox="1559 746 1655 786"></th> <th data-bbox="1655 746 2040 786">Evidence</th> </tr> </thead> <tbody> <tr> <td data-bbox="1290 786 1559 866">1.1 Goal setting (behaviour)</td> <td data-bbox="1559 786 1655 866">++</td> <td data-bbox="1655 786 2040 866">The goal is set for 30 min session</td> </tr> <tr> <td data-bbox="1290 866 1559 970">1.4 Action planning</td> <td data-bbox="1559 866 1655 970">++</td> <td data-bbox="1655 866 2040 970">Action plan includes 3 sessions per week for 9 weeks</td> </tr> <tr> <td data-bbox="1290 970 1559 1074">2.2 Feedback on behaviour</td> <td data-bbox="1559 970 1655 1074">+</td> <td data-bbox="1655 970 2040 1074">Audio couch informs when each running session is complete</td> </tr> <tr> <td data-bbox="1290 1074 1559 1209">4.1 Instruction on how to perform the behaviour</td> <td data-bbox="1559 1074 1655 1209">++</td> <td data-bbox="1655 1074 2040 1209">Audio coach instructions while walking/running; written tips and advice section</td> </tr> <tr> <td data-bbox="1290 1209 1559 1321">5.1 Information about health consequences</td> <td data-bbox="1559 1209 1655 1321">++</td> <td data-bbox="1655 1209 2040 1321">Tips and advice section includes a description of the health benefits</td> </tr> </tbody> </table>	BCT name		Evidence	1.1 Goal setting (behaviour)	++	The goal is set for 30 min session	1.4 Action planning	++	Action plan includes 3 sessions per week for 9 weeks	2.2 Feedback on behaviour	+	Audio couch informs when each running session is complete	4.1 Instruction on how to perform the behaviour	++	Audio coach instructions while walking/running; written tips and advice section	5.1 Information about health consequences	++	Tips and advice section includes a description of the health benefits
BCT name		Evidence																																				
1.1 Goal setting	++	The exercises are defined and set for the user																																				
2.2 Feedback on behaviour	++	The calendar records the exercises done																																				
2.4 Self-monitoring of outcome(s) of behaviour	++	The option to record and track the weight changes																																				
2.7 Feedback on outcome(s) of behaviour	++	The graph shows weight change (inputted by the user)																																				
4.1 Instruction on how to perform the behaviour	++	The exercises are explained in text and audio																																				
BCT name		Evidence																																				
1.1 Goal setting (behaviour)	++	The goal is set for 30 min session																																				
1.4 Action planning	++	Action plan includes 3 sessions per week for 9 weeks																																				
2.2 Feedback on behaviour	+	Audio couch informs when each running session is complete																																				
4.1 Instruction on how to perform the behaviour	++	Audio coach instructions while walking/running; written tips and advice section																																				
5.1 Information about health consequences	++	Tips and advice section includes a description of the health benefits																																				

<p>Example screenshots of the circuit session</p>	5.1 Information about health consequences	++	The introduction details some possible health benefits	5.6 Information about emotional consequences	++	Tips and advice section includes a description of the benefits “to the mind”
	6.1 Demonstration of the behaviour	++	The exercises are shown as videos and graphs	6.2 Social comparison	+	Tips and advice section includes a “success stories”
	7.1 Prompts/cues	++	Reminders to perform the PA	9.1 Credible source	++	App developed by PHE
	9.1 Credible source	++	The introduction specifies that the app is based on reviewed scientific research with a link to the study	10.4 Social reward	++	Verbal encouragement from the audio coach after finishing the session
	10.4 Social reward	++	The audio provides encouragement and praise, “Workout complete. Congratulations”	2.2 Feedback on behaviour	+	Audio coach informs when each running session is complete
	10.1 Material incentive (behaviour)	++	The bonus packs of exercise that costs money would be free if behaviour was performed (this is measured through the achievements)	10.5 Social incentive		Audio coach encourages before the running session, e.g., “You can do it!”
				12.1 Restructuring the physical environment	++	Audio coach advices of the possible changes to the environment if the exercise gets monotonous: “it’s 7th week and the running might get a bit monotonous. There are a few things you could do. Maybe if you run on a treadmill try to run outside or if you already run outside, try to change the route”

		<p>12.2 Restructuring the social environment</p>	<p>+</p>	<p>The audio coach says: "Also, you may try to find a buddy to run with"</p>
		<p>Example screenshot of the running session</p> 		
<p>4. What – procedures</p>	<p>Through video and audio instructions the user is instructed which and for how long to perform the exercises during the workout</p>	<p>Through audio instructions the user is instructed when to walk and when to run during the workout</p>		
<p>5. Who provided</p>	<p>Fitness Guide Inc.</p>	<p>PHE</p>		
<p>6. How</p>	<p>Fully automated app-based</p>	<p>Fully automated app-based</p>		
<p>7. Where</p>	<p>The workout targeted in the app can be performed anywhere with an access to a wall and raised surface as the workout uses body weight</p>	<p>As it involves alternate walking and running, the workout session will most likely be performed outdoors</p>		
<p>8. When and how much</p>	<p>The app instruction does not specify the time and frequency of the workout. The instruction within the app</p>	<p>The app programme is spaced through 9 weeks and 3 sessions per week are recommended. Each session lasts 30 min</p>		

	suggests to conduct the 7 min circuit “2-3 times if time allows”	
9. Tailoring	The workout can be tailored by changing the duration of the exercises and the rest time, by customising the order of the circuit and by setting the order of exercise to “random”	The voice of the coach can be chosen from 5 options of public figures
10. Modifications	The apps were monitored during the duration of the study and no changes were made to the features of the apps	
11. How well – planned	The instructions provided to the participants were to use the app in a self-directed way	
12. How well – actual	The intervention usage was one of the outcomes of the study	

### **4.3.5 Procedures**

#### **Recruitment**

Participants were recruited through posters displayed in public areas and through the website [callforparticipants.com](http://callforparticipants.com), and social media. Interested participants accessed the study website (managed using Qualtrics) through the link provided in the study advertisements (See Appendix L for the study material). The website included the participant information sheet (PIS), the screening questionnaire, and the consent form. Participants who fulfilled the inclusion criteria were asked to provide electronic written consent. The researcher contacted the study participants immediately after receiving the consent to schedule the baseline assessment.

#### **Data collection**

Data were collected face-to-face at baseline. Post-baseline (week 1), and follow ups at 3 and 5 weeks were conducted online using Qualtrics. Participants used the accelerometer for 7 days to assess the baseline PA level, and then at week 3, and 5.

#### **Randomisation**

Following completion of the baseline 1 week assessment of the accelerometer-measured PA and post-baseline questionnaire, participants were randomised to one of the two app assessment sequences (App A first, then B or App B first, then A). The randomization code was developed using a computer random number generator to select random permuted blocks of size 4. The investigator requested each randomisation assignment at the time of randomisation and did not have access to the whole list managed by AS.

Each app was used by participants for two weeks each. Participants received a short instruction with a link to the first app (based on randomisation sequence) and were asked



to use the apps in a self-directed way with an aim to increase their PA level. No instruction of the frequency of usage was provided as this trial aimed to mimic the “real-world” conditions.

At the end of the study, participants received £20 as an incentive for participation and to encourage completion of the study measures. In addition, each participant was reimbursed £3 for downloading of app A. The provision of the reward was not dependent on the usage of the app as exploring the adoption of the app into the users’ daily routines is one of the objectives of the qualitative component of this study (phase 2, study 4).

#### **4.3.6 Measures**

Socio-demographic characteristics of age, gender (male/female), ethnicity, country of birth and years lived in the UK, relationship status, highest education level, employment status, and household income were measured. These particular measures were used as they have been shown to be determinants / correlates of PA [263] (Appendix J).

The participant’s main motivation for attempting to increase PA was measured with one question: ‘What is the main reason you want to increase your PA? Appearance (e.g. improve body shape); Competence (e.g. obtain new skills, to challenge oneself); Fitness (e.g. improve physical health, well-being); Other, specify; I don’t want to increase my PA.’ The goal of this question was to assess the source of the motivation (extrinsic/intrinsic) as there is evidence of different outcomes for these two types of motivation [246, 247]. As no one-question assessment was found in the research

literature, this question was designed by the investigator to assess the baseline difference in the motivation between the groups.

#### **4.3.6.1 Open-ended questions**

Two open-ended questions were used in the study. The first was dependent on whether the participants indicated they had downloaded apps for PA before. When answered in the affirmative, the open-ended questions asked what kind of PA was targeted in the previous apps. The second question asked about the expectations from the app interventions.

#### **4.3.6.2 Objective PA assessment using accelerometer**

Participants wore a GT3X+ accelerometer (actiGraph, Pensacola, Florida) on an elasticated belt over the right hip for 21 days (7 days each at baseline, week 3, week 5). Participants were asked to wear the device during their waking hours except for water-based activities.

Daily PA count was used as the primary outcome to measure the effects of apps on PA. It was operationalised as Axis 1 Counts Per Minute (CPM) for Axis 1 (Y-Axis) in the Actilife software. This measure is calculated by summing all Axis 1 Counts and then dividing by the total number of minutes recorded by the device. This value can be interpreted as the average counts for each minute on Axis 1.

In addition, as this was a feasibility trial exploring the most appropriate measures for the definitive trial, other PA outcomes were explored, specifically: MVPA, light, moderate, vigorous, sedentary behaviour, step count, and 20% increase in MVPA.

### ***Wear and non-wear time definitions***

In this study, the wear time of a minimum of 480 min (at least 8 h per day), for at least 3 days was required. This method was used in previous studies [282-285]. Non-wear time was defined as a continuous string of zeros for >90 min, with *artifactual movement tolerance* of 2 min (small spikes of non-zero activity lasting up to 2 min within the 90 min period). If 30 min before and after the spike showed consecutive zeros within those 90 min, this period was considered as non-wear time. This algorithm is designed to accommodate any accidental movement (*artifactual movement*) of the device. This definition has been validated and widely used in PA research [286].

### ***Epoch length***

There are no data on the influence of the epoch length (i.e., accelerometer time-sampling intervals) on the accuracy of accelerometry output data in adult population. In children, Aibar, Bois [287] found that smaller epoch lengths had influence on the granularity of the measure, suggesting the use of shorter (3-15 seconds) epoch lengths to make the measure more accurate. For this study, the raw data collected were integrated into 5s epochs.

### ***Data processing***

Data were processed using Actilife software (version 6.13.3, actiGraph, LLC). Freedson's cut-off points [288] ,i.e., the thresholds that categorise the CPM into PA intensities were used to define time spent in sedentary (0 -99 CPM), light (100 - 1951 CPM), moderate (1952 - 5724 CPM), Vigorous: (5725 - 9498 CPM).

The secondary outcomes were self-reported PA, psychological outcomes: ESE, intentions, PA outcome expectancy and mood.

#### **4.3.6.3 Self-reported physical activity**

Self-reported PA was assessed using the International Physical Activity Questionnaire (IPAQ) Short Form [261] which includes 9 items assessing walking, moderate, vigorous PA, and SB.

The score for IPAQ is expressed in METs. MET-minutes are computed by multiplying the METs of specific behaviours by the minutes performed (METs are described in the Introduction Chapter). The IPAQ processing guidelines were followed to generate total time spent in PA, SB (min/week), moderate, vigorous, walking and total PA (in MET-min/week) [289].

#### **4.3.6.4 Psychological measures: Exercise self-efficacy, intentions, PA outcome expectancy**

##### **Outcome expectancy (OE)**

The expected outcomes were measured using the Expected Outcomes for Physical Activity scale [290] which includes 12 items asking participant to record their level of agreement with the statements on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). Higher average scores indicate higher perceived benefits of PA.

##### **Exercise intentions**

Exercise intention was measured by two items using a 7-point scale as recommended by Ajzen (2002) from 1 (strongly disagree) to 7 (strongly agree) with higher average scores indicating stronger intention to perform PA.

##### **Exercise self-efficacy (ESE)**

Self-efficacy was measured using the Self-Efficacy to Regulate Exercise questionnaire which included 18 items asking participants about their confidence in conducting PA in various contexts. The measures range from 0% to 100% confidence with a 10-unit

interval [291]. The average for the measure is calculated and higher scores indicate higher confidence in conducting PA.

## **Mood**

There is a bi-directional relationship between PA and affective states [292, 293]. In addition, mood is a state rather than a trait; it is prone to fluctuation [294-296]. Hence, to consider the impact of mood variations on PA, in this study, the mood state was measured using the Ecological Momentary Assessment (EMA). This instrument conducts real-time data collection in the natural environment of the participants. Mood during the study was assessed to investigate the acceptability of using such measure, to assess the average mood throughout the study, and to assess the relationship between the physical activity and mood.

PACO by Google, an open source app-based software was used to collect mood state. Signal-contingent sampling was used where prompts were sent to participants at random times between 9am and 20:15pm (unless the participant specified otherwise) by asking the question: "How is your mood right now?" and recording the response using visuals of facial expressions from sad (1) to happy (5). Data were collected during 3 weeks of the 5 week trial (baseline, week 3, week 5). In total there were 42 assessments for each participant (21 days x 2 per day).

## **Usability**

The measure used to operationalise usability was the SUS [161] which consists of 10 items that are ranked on a 5-point Likert scale which yields a score from 0-100. This measure was used in Study 1 of this thesis as one of the measures to assess the quality of user experience of popular PA apps (See section 'SUS', Chapter 2)

## **User ratings**

Participants were asked to award 1 to 5 stars to the apps to mimic the app stores ratings. This variable was extracted and used to characterise user experience in Study 1 (See section 'User ratings', Chapter 2).

## **Engagement**

Engagement with apps was assessed using the digital behaviour change intervention (DBCI) Engagement Scale [272]. The measure consists of 3 components assessing *experiential items* scored on a scale from 1-7, the *amount of use* asking the participants to report the time spent on using the apps (in min), and *depth of use* which is reported as the proportion of the app components used by the participants. In this study, there was a maximum of 8 components for App A (Introduction, Exercise list, Begin workout, Activity calendar, Results tracker, Achievements, Extra workouts, Reminders) and 5 for App B (My runs, Choice of coach, Reminder option, Support section, Additional resources). In addition, a total engagement score can be assessed combining these 3 components. However, the time to complete one session for App A and B differed (7 min versus 30 min, for App A and B respectively) and hence, the apps were not comparable, the total score for app engagement was not assessed.

### 4.3.6.4.1 Data gathered to facilitate the interviews (Study 4)

## **App usage**

Participants were asked to send the screenshots of the app feature that showed the completed PA sessions (*Activity calendar* for App A and *My runs* for App B) to ensure that the app was downloaded and so that the researcher could see the engagement with the app. These screenshots were utilised to prompt participants during the data-prompted interviews.

In addition, diaries were used to record any experiences of adapting the app into the users' daily routines (various media: text, photography, audio and video recordings). These diaries were used to explore the experiences of using PA apps in the qualitative component of this study (phase 2).

Table 40 below summarises the measures used in the study specifying the time point of the assessment.

Table 40: Summary of data collection at different data collection points

	Baseline	Post-baseline	Week 3 and 5
Socio-demographic data			
Age	●		
Gender	●		
Ethnicity	●		
Relationship status	●		
Highest education level	●		
Employment status	●		
Household income	●		
PA			
actiGraph GT3X+	21 days (week 1, 3, 5)		
International Physical Activity Questionnaire	●	●	●
Psychological outcomes			
Exercise intentions	●	●	●
Expected Outcomes for Physical Activity	●	●	●
Exercise self-efficacy	●	●	●

Ecological Momentary Assessment of mood	Twice a day throughout the study		
App-specific measures	●	●	●
System Usability Scale			●
Engagement Scale			●
Experience of using the apps			
Multimedia diary data	ad hoc, at participants' convenience		

#### 4.3.6.4.2 Assessment of the stability of the pre-intervention baseline

The differences between the baseline and post-baseline self-reported outcomes were assessed to determine if the relevant variables were already changing as an effect of merely taking part in a study. The act of answering questions about a behaviour, or knowing that the behaviour is being recorded by an accelerometer may change the behaviour without the introduction of the intervention, known as the mere measurement effect, reactivity of assessment [297, 298] or Hawthorne effect [299].

### 4.3.7 Outcomes

The outcomes to assess feasibility, acceptability, and the potential effects of the apps are listed below.

#### 4.3.7.1 Feasibility and acceptability

1. Recruitment rates as a percentage of eligible participants (i.e., completed screening questionnaire and signed the electronic consent form), number of participants enrolled and randomised (met with the researcher for the baseline assessment)



2. Retention measured by adherence to wearing the accelerometer on at least 3/7 days of each data collection period and completeness of follow-up questionnaire data at post-baseline, 3 and 5 weeks' follow up
3. Acceptability of the trial procedures (randomising participants to apps, data collection methods) and the interventions was also assessed in the semi-structured interviews following the completion of the trial (Study 4)

#### **4.3.7.2 Potential effects of the interventions on behavioural and psychological outcomes, and usability of and engagement with the apps**

The outcomes to assess the potential effects of the apps were:

1. Change from baseline to 3 weeks follow up in objectively measured PA: daily PA count (vertical count acceleration, CPM), MVPA, light, moderate, vigorous, SB, step count
2. The proportion of participants that increased their time in MVPA by 20% from baseline
3. Change from baseline to 3 weeks follow up in self-reported:
  - PA
  - Expected outcomes
  - Exercise intentions
  - ESE
  - Mood
4. The difference in change between the two apps
5. Usability and user ratings of the apps
6. Engagement with the apps

The primary outcome to assess the potential effects of the apps on behaviour was daily PA count, and was measured using accelerometer.

A continuous outcome was selected because it was the most powerful to represent the potential population impact. In addition, there is evidence that even a small increase in PA is beneficial to health [57]. As this was a feasibility study and given there was likely to be substantial variability in PA outcome, it was of interest to also explore a categorical outcome. 20% increase in MVPA was a pragmatic choice based on a conversation with PA expert (B. Jefferis, personal communication, September 11, 2017) exploring the options for a meaningful dichotomous outcome.

#### **4.3.7.3 Sensitivity analyses**

Analysis was conducted including those participants that engaged with the app intervention defined as those below and above the median measured using the DBCI Engagement Scale variable *amount of use* of the app in minutes.

Analysis of the potential influence of weather on PA was investigated. Specifically, the potential impact of the temperature, rain and snowfall were assessed.

Axis 1 CPM was chosen as, in most of the literature using accelerometer data, the y-axis (axis 1, vertical axis pointing up) was predominantly used to capture PA. This is because the earlier generation of devices had the technology to capture data on only one axis. However, in more recent years the technology enables the use of triaxial accelerometers for which the outcome is vector magnitude, and hence sensitivity analysis was run on the vector magnitude CPM.

Lastly, post-hoc sensitivity analyses were conducted based on the results. These were not planned but conducted as needed. These included the log transformation of the IPAQ measure due to the skew of several PA variables and outliers in the OE measurement.

#### **4.3.8 Ethical approval**

This protocol and related documents were submitted for review to University College London's Research Ethics Committee. Following the UK Data Protection Act 1998, any documents containing participants' identifiable information were kept in a locked cabinet at the department of Primary Care and Population Health, UCL. Those documents were kept separate from other research data in order to ensure confidentiality and information security.

In addition, audio recordings of the interviews (Study 4) intentionally did not record participants' names and personal information. Audio files were given a code number not linking them to the names of the participants and they were kept in UCL electronic Data Safe Haven. The data were transcribed by a professional transcriber. All data were anonymised, and no identifying data were reported. Transcripts were kept in a secure location and only I had access to these files.

#### **4.3.9 Summary of Patient and Public Involvement**

The patient and public involvement (PPI) representatives included a 50-year-old woman, and a 55-year-old man. Both resided in London and engaged in little PA. They reviewed the proposed design of the study. They reviewed the study materials (PIS, online study advertisements). They also provided some valuable feedback on the topic guide development for the qualitative study 4 to ensure the questions were relevant, and the language used was comprehensible and jargon-free. They also assisted with the interpretation of the results.

#### **4.3.10 Sample size**

This study was not powered to detect a difference in effectiveness between the interventions. As the primary goal of the study was feasibility, the sample size of 60 participants was selected based on resource considerations and formal sample size calculations were not undertaken to inform this choice. The National Institute for Health recommends a sample size of 30 in each arm as a pragmatic rule [300] for feasibility studies, and this recommendation was the main factor in selecting the sample size.

#### **4.3.11 Analysis**

The analyses were performed using the ITT population. While this study provides statistical analysis of the endpoints collected, there was no formal hypothesis testing, and so no type 1 error allocated in the design of the study for inferential analysis (to control the rate of false-positive results). In consequence, these results are intended for hypothesis generation and to support planning for future studies. No statistical inference should be drawn from these findings, and statistical tests (p-values) should be interpreted as descriptive.

Descriptive statistics were used to report the socio-demographics and other characteristics of the participants, recruitment, retention rates, mood, and app-specific characteristics: usability, user ratings and engagement. Student's t-test for independent samples, the Wilcoxon Rank-Sum test and the chi-squared or Fisher's exact test were used to compare baseline characteristics and other single-participant measures.

The paired Student's t-test was used to compare the difference between continuous and relatively normally distributed variables: usability for App A and B, star ratings and engagement.

Wilcoxon Signed Rank test was used to assess the difference in the baseline and post-baseline (week 1), and the difference in PA count change for high versus low mood.

The two open-ended question responses were re-coded according to the similarity of the content to identify themes in the responses. Following this classification, the themes and frequency of the responses were generated using NVivo.

Recent randomised crossover trials in behavioural interventions were used to guide the analysis approach [301]. The primary endpoint for the study was the change from baseline in the objective measure of PA (accelerometer). The primary effectiveness analysis compared the difference in the intra-participant changes from baseline to the period in which the participant was using the first assigned app (Period 1). This endpoint was assessed using a mixed model for repeated measures including fixed effects for period and baseline activity, and a random effect for app (App A or App B). This is similar to an ANOVA model that accounts for two different measurements (one from each period) for the same participant.

The proportions and CIs for patients achieving 20% or greater increase in MVPA were calculated.

A secondary question of interest was whether there was any difference in change in PA across the apps. This endpoint was analysed using a mixed model for repeated measures including baseline PA, app (App A or App B), period, and the app x period interaction as fixed effects.

As a sensitivity analysis to assess the relationship between PA and weather, a mixed model for repeated measures was run including fixed effect for app and weather

variables (any rain and any snow as binary variables and continuous mean temperature) with daily PA count as the dependent variable. The app-by-weather interactions for rain and snow were also tested to determine whether the relationship between app and daily PA was impacted by weather.

In conclusion, in this Chapter I described the processes of selecting the study design and methods. In the next Chapter I describe the results and discussion of this crossover feasibility trial of two apps.

## **CHAPTER 5. Results and discussion: the feasibility and acceptability of a randomised crossover trial assessing two popular physical activity apps**

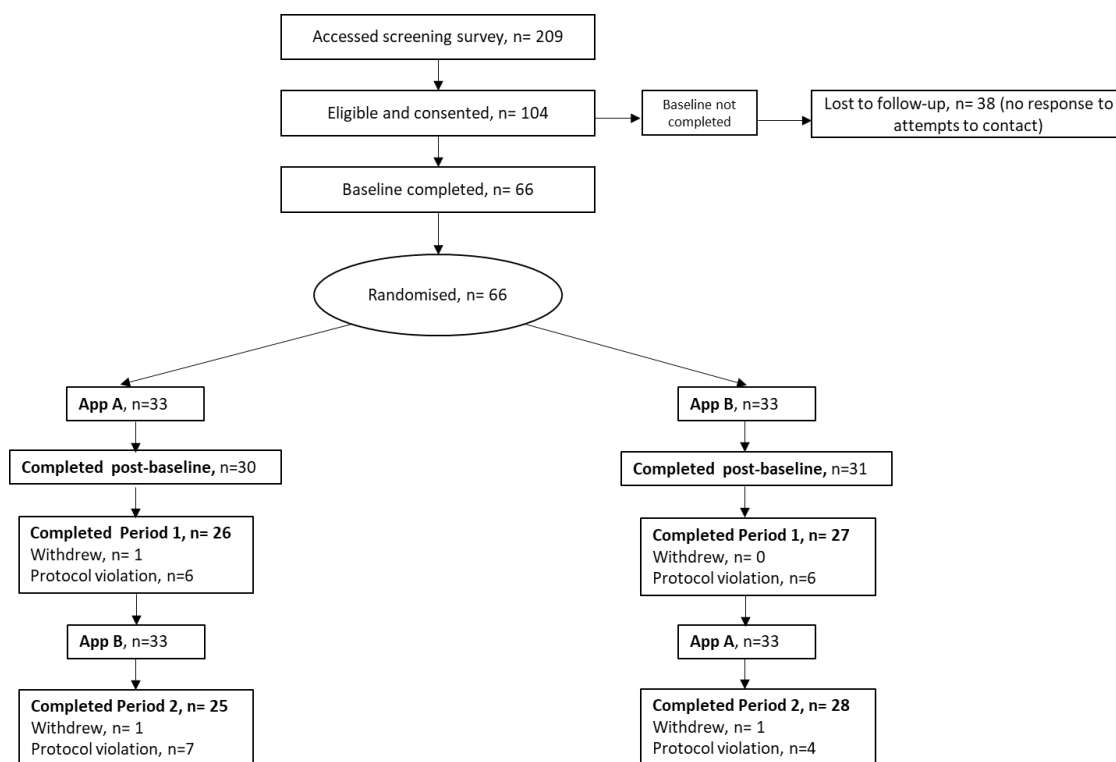
### **5.1 Chapter overview**

This chapter reports on the results and discusses the findings of the quantitative component of this feasibility crossover trial of two popular apps on the market.

### **5.2 Results**

Participants were consented between 15th of January and 13th of April 2018 with the final follow-up conducted on 19 of May. See Figure 39 for CONSORT participant flow diagram [277]. The recruitment ended after 66 participants were enrolled to the study and completed the baseline assessment.

Figure 39: Flowchart for recruitment and retention to the crossover trial



*Adapted from the CONSORT guidelines Schulz, Altman [302]*

*Note:* Instances of protocol violation were: 10 participants did not adhere to the accelerometer wear protocol, 1 participant did not download the app.

### 5.2.1 Baseline characteristics

Participant characteristics at baseline are presented in Table 41. The mean age was 31.1 years, and 42/66 (63.6%) were women. Twenty three of 66 (34.8 %) described themselves as non-White and 48 (72.7 %) were born in the UK. The majority of participants were single (39/66, 59.1%). Less than half of participants completed postgraduate education (29/66, 43.9%). Twenty eight of 66 of participants were in full time employment (42.4%) and 19/66 (28.8%) were in full time education. Two participants were retired, 2 were self-employed, and 4 were unemployed. The most frequently reported range of monthly household income was £1,001 - £3,000 reported by 30/66 participants (45.5%). The majority of participants (65.2%) had downloaded a



PA app before. Most of the participants had not downloaded running programme-type apps or HIIT-type apps before (58/66, 87.9% and 63/66, 95.5%, respectively). Fifty one of 66 (77.3%) had never used a wearable device to measure their PA. Of 15 participants who reported previous use of a wearable device, 9 reported using it regularly. None of the participants were taking part in another study. Fitness was the most common main motivator with 44/66 (66.7%) participants reporting increasing their fitness as their main motivator for attempting to increase PA. The second most common motivator was appearance (29/66, 43.9%). Competence was not frequently reported as a motivator (8/66, 12.1%). One participant mentioned another motivator which was to “have better well-being”.

Table 41: Baseline characteristics of the sample (n=66)

### Baseline characteristics

		App A first (n=33)	App B first (n=33)	All (n=66)
Age (years)	Mean ± SD	29.5 ± 10.4	32.7 ± 12.1	31.1 (11.4)
Gender (N,%)	Female	22 (66.7)	20 (60.6)	42 (63.6)
	Male	11 (33.3)	13 (39.4)	24 (36.4)
Ethnicity (N,%)	Asian	5 (15.2)	8 (24.2)	13 (19.7)
	Black	3 (9.1)	0 (0.0)	3 (4.6)
	Mixed	2 (6.1)	3 (9.1)	5 (7.6)
	White	22 (66.7)	21 (63.6)	43 (65.2)
	Other	1 (3.0)	1 (3.0)	2 (3.0)
Duration in the UK (years)	Mean ± SD	6.67 ± 3.79	10.1 ± 13.5	3.5 (7.8)
Relationship (N,%)	Single	20 (60.6)	19 (57.6)	39 (59.1)
	In a relationship	13 (39.4)	12 (36.4)	25 (37.9)
	Separated	0 (0.0)	2 (6.1)	2 (3.0)

## Baseline characteristics

		App A first (n=33)	App B first (n=33)	All (n=66)
Education (N,%)	Postgraduate	15(42.4)	15 (45.4)	29 (43.9)
	Undergraduate	12 (36.4)	7 (21.2)	19 (28.8)
	Primary/secondary/college	7 (21.2)	11 (33.3)	18 (27.3)
Occupation (N,%)	Full-time education	9 (27.3)	10 (30.3)	19 (28.8)
	Full-time employment	16 (48.5)	12 (36.4)	28 (42.4)
	Part-time employment	3 (9.1)	4 (12.1)	7 (10.6)
	Retired	1 (3.0)	1 (3.0)	2 (3.0)
	Self-employed	0 (0.0)	2 (6.1)	2 (3.0)
	Unemployed	2 (6.1)	2 (6.1)	4 (6.1)
	Other	2 (6.1)	2 (6.1)	4 (6.1)
Household income (monthly, N,%)	Under £1,000	2 (6.1)	4 (12.1)	6 (9.1)
	£1,001 - £3,000	17 (51.5)	13 (39.4)	30 (45.5)
	>£3,000	11 (33.3)	11 (33.3)	22 (33.3)
	Not applicable	3 (9.1)	5 (15.2)	8 (12.1)
Downloaded PA apps before (N,%)	Yes	21 (63.6)	22 (66.7)	43 (65.2)
	No	12 (36.4)	11 (33.3)	23 (34.9)
Nr of apps downloaded before (N,%)	0	12 (36.4)	11 (33.3)	23 (34.9)
	1	10 (30.3)	11 (33.3)	21 (31.8)
	2	7 (21.2)	5 (15.2)	12 (18.2)
	3	2 (6.1)	4 (12.1)	6 (9.1)
	≥4	2 (6.1)	2 (6.1)	4 (6.1)
Downloaded running programme-type app before (N,%)	Yes	5 (15.2)	3 (9.1)	8 (12.1)

### Baseline characteristics

		App A first (n=33)	App B first (n=33)	All (n=66)
	No	28 (84.8)	30 (90.9)	58 (87.9)
Downloaded HIIT-type app before (N,%)	Yes	1 (3.0)	2 (6.1)	3 (4.6)
	No	32 (97.0)	31 (93.9)	63 (95.5)
Used wearables before (N,%)	Yes	5 (15.2)	10 (30.3)	15 (22.7)
	No	28 (84.8)	23 (69.7)	51 (77.3)
Use wearable regularly (N,%)	Yes	3 (9.1)	6 (18.2)	9 (86.4)
	No	30 (90.9)	27 (81.8)	57 (13.6)
Taking part in another study (N,%)	No	33 (100.0)	33 (100.0)	66 33 (100.0)
Main motivators for increasing PA (N,%):	Appearance			
	Yes	18 (54.6)	11(33.3)	29(43.9)
	No	15(45.5)	22(66.7)	37(56.1)
	Competence			
	Yes	28 (84.9)	3 (9.1)	8(12.1)
	No	5 (15.2)	30 (90.9)	58 (87.9)
	Fitness			
	Yes	18 (57.6)	25 (75.8)	44 (66.7)
	No	14 (42.4)	8 (24.2)	22 (33.3)
	Other			
	Yes	1(3)	0	1 (1.5)
	No	32 (97.0)	33 (100.0)	65 (98.5)

### 5.2.1.1 Expectations from the app interventions

The word cloud of the original participants' responses to the open-ended question (the more frequent words used to describe expectations represent larger font) is presented in Figure 40. The responses were re-coded according to the similarity of their content to establish the frequency of responses. For example, participants' responses "to improve my physical condition" and "I can get fit" were re-coded to "improve fitness"; responses: "keeping data of workout" and "data (steps, calories, heart rate)" were re-coded "to monitor the activity"; and response "some guidance, a framework and overview of past activity" was re-coded into 3 expectations: 1. guide, 2. track activity, 3. feedback on activity.

Subsequently, participants' responses were grouped into themes. Table 42 presents the themes from the open-ended question asking about the expectations from the app interventions. The themes that emerged from the data referred to what the participants expected the app to do (the "what") and how the apps were expected to help with the PA (the "how").

#### 5.2.1.1.1 Expectation of the "what"

The themes that emerged relating to the expectations of what the apps would provide were described using the stages of the process of performing PA. Firstly, before the behaviour can occur, there needs to be a motivation to engage in PA. Increasing motivation was the most common response with 26 participants expecting the apps to motivate them to do PA (as visualised in Figure 40). The frequency of the motivation expected from participants taking part in this study is manifested in the word cloud of the frequency of responses (Figure 40). Secondly, the content of the app, specifically the provision of PA guidance, knowledge and expertise was reported as a frequent expectation from the app. Third, after increase in motivation and provision of content, establishing a routine and setting goals was an expectation reported frequently, and

fourthly, once a routine is set, help with the adherence to this routine was expected from the apps. Lowest in frequency were participants' responses related to their various anticipations to achieve an increase in PA (and/or reach targets and goals), and achieve outcomes (e.g., lose weight, improve health, increase energy).

#### 5.2.1.1.2 Expectation of the “how”

Three themes emerged relating to the expectations of how the apps would address PA. First, participants expected the apps to track (monitor, feedback) the activity itself and the outcomes of the activity (e.g., fitness, calories burnt, and achievements). Second, there were expectations that there would be reminders in place to prompt the participants to engage in PA. Lastly, some participants expected some digital rewards for their progress.

Figure 40: Word cloud representing the most frequently words used by the participants to describe their expectations of the app interventions.



Table 42: The 8 themes that emerged to the responses to the open-ended question:

What do you expect from the apps?

Themes	Re-coded responses clustered under the theme
<b>The what</b>	
1. Motivate	motivate (26), to encourage me to do exercise (3), not give up so easily, help with self-discipline
2. Provide guidance/knowledge/expertise	provide information/ knowledge/expertise (4), guide (3), new exercise ideas, original and new stuff, give me tips
3. Establish activity routine/goals	set goals/targets (5), set a routine (4), exercise plan, to help in structuring fitness activity
4. Achieve an increase in activity/achieve goals	increase exercise (8), achieve goals, goals/targets
5. Achieve outcomes	lose weight (3), Improve/boost fitness (3), improve/increase health (2), get fit (2) overall well-being, increase energy, increase strength
6. Adhere to a routine/goals	to adhere to an activity routine (3), to build a habit, regular exercise
<b>The how</b>	
7. Tracking of activity/outcomes of activity	track activity (13), to track progress (5), to monitor activity (3), to feedback on activity (2), to track calories burnt, monitor fitness level, monitor exercise level, to track achievements, to track diet, provide feedback on goal achievement
8. Reminders	to remind (5), to prompt (2), reminder if I am underperforming
9. Rewards	to reward (4)

### 5.2.1.2 Previous PA app usage

Forty-three out of 66 (65.2%) participants reported that they had downloaded PA apps previously.

The responses to the open-ended question asking participants what PA was targeted in the apps they have downloaded previously were first re-coded according to the similarity of the content to establish the frequency of responses. Table 43 presents the type and the frequency of the PA apps downloaded prior to taking part in the trial.

The most common type of PA app reported were those that targeted running (24 apps), followed by a workout (14), and walking. General activity trackers were downloaded 4 times in the sample of participants in this study. Within the “Other” category, 1 app was event based: “outside walking/running which raised money for charity”, and another targeted weight loss but the PA targeted was not reported.

Table 43: The type and the frequency of the PA apps downloaded prior to taking part in the trial

<b>PA targeted</b>	<b>Nr of responses</b>
Running	24
Workout	14
Walking	9
Activity tracking	4
Yoga	3
Gym-based app	3
Pilates	2
Weight lifting	2
Cycling	2
Other	2
Total number of PA apps downloaded	65

## **5.2.2 Outcomes**

### **5.2.2.1 Feasibility and acceptability**

The qualitative phase of the trial exploring acceptability of the trial procedures and the interventions (randomising participants to apps, data collection methods) is presented in Study 4 (section 'Acceptability of the trial design and procedures', Chapter 6).

#### **5.2.2.1.1 Recruitment and retention rates**

##### ***Recruitment rates***

Two hundred nine people accessed the screening questionnaire. Out of 104 participants who were eligible and consented, 63.5% (66/104) were enrolled and randomised.

##### ***Retention in the trial***

##### **Completeness of accelerometer data**

Over three-quarters (52/66; 78.8%) of participants provided valid accelerometer data ( $\geq 3$  days of data with  $\geq 480$  min each day) for the baseline and at least one of the intervention periods. Eleven participants provided no 2<sup>nd</sup> app period data, 1 had no 1<sup>st</sup> app assessment data, and 14 participants did not provide sufficient data. Of these 14 participants, 2 were cases where participants were excluded from accelerometer analyses after randomisation as their device was not compatible with app B.

##### **Completeness of follow up questionnaire data**

The post-baseline questionnaire was completed by 61/66 (92.4%) of participants. For 51/66 (77.3%) of participants complete 3 weeks follow up data were available, and 52/66 (78.8%) of data were available for 5 week follow up.



The qualitative phase of the trial exploring the acceptability of the trial procedures and the interventions (randomising participants to apps, data collection methods) is presented in Study 4 (section 'Acceptability of the trial design and procedures', Chapter 6).

### **5.2.2.2 The effects of the apps on objectively measured PA, self-reported PA, and psychological outcomes**

#### **5.2.2.2.1 The stability of the pre-intervention baseline**

Due to technical issues with the Qualtrics software, two items in the questionnaire were not recorded correctly: the 12<sup>th</sup> item in the OE scale for the baseline assessment and item 18 on the ESE scale for the follow up assessments. The convention in patient reported outcomes is that a score can be calculated when greater than 50% of items are non-missing [303, 304]. Although the manuals for the instruments used in this study do not explicitly give guidelines for missing data, applying this rule was preferable to not performing the analysis. Because the missingness was administrative, this is a rare case where the missing data are guaranteed to be un-related to the responses. Hence, to analyse the items, for OE, a new score was calculated defined as the sum of the first 11 items and this score was used to compare change from baseline. Similarly, for ESE, the average of the 17 items was used.

Wilcoxon Signed Rank test for median differences in median baseline and post-baseline distributions in self-reported PA showed that participants reported more vigorous PA at post-baseline ( $p < 0.001$ ), higher Intentions ( $p < 0.001$ ) and lower OE ( $p < 0.001$ ). The difference in OE was sustained after removing one extreme outlier in which the change score was +41 ( $p < 0.001$ ). No other tests of difference in PA measures or ESE were found to be significant.

Further inspection of the histograms and the descriptive statistics of the significant results showed that median and mean difference in vigorous PA (MET-min/week) between baseline and post-baseline was 0. However, the difference in distribution was significant. Seventy percent of the participants had changes between 480 and -480 METs. Ten percent of the participants had changes between 960 and 1920 which is why the test was statistically significant. An equivalent of 3000 MET minutes each week can be achieved by climbing the stairs for 10 minutes or running for 20 minutes on a daily basis [305].

#### 5.2.2.2.2 Objective outcome: accelerometer

##### **Change from baseline to 3 weeks follow up in objectively measured PA**

The mean daily wear time for participants with valid data for baseline was 801.84 min (SD 84.12), for Period 1: 784.62 min (SD 89.63), and for Period 2: 819.76 min (SD 99.22).

The primary analysis of the accelerometer data showed that there were no significant differences between baseline and Period 1 change on any of the objective PA outcomes. (Table 44).

Table 44: Primary analysis – objective measure of change in PA (accelerometer): change in daily PA count (Axis 1 CPM), change in MVPA, time spent in light, moderate, vigorous PA, SB, step count, and wear time from baseline to Period 1 (in minutes)

	Period 1			p-value	Period 2			p-value
	LSM	95% CI lower	upper		LSM	95% CI lower	upper	
PA count	10.57	-22.32	43.46	0.525	-2.69	-39.37	34.00	0.885
MVPA	0.96	-4.43	6.36	0.723	0.91	-5.11	6.93	0.764
Light	-5.06	-11.76	1.64	0.137	-0.08	-7.55	7.40	0.983
Moderate	0.24	-4.81	5.30	0.924	1.12	-4.52	6.75	0.695

Vigorous	0.67	-0.68	2.02	0.329	-0.14	-1.65	1.37	0.855
SB	-13.81	-36.57	8.95	0.231	17.46	-7.93	42.84	0.175
Step count	10.57	-22.32	43.46	0.525	-2.69	-39.37	34.00	0.885
Wear time	0.96	-4.43	6.36	0.723	0.91	-5.11	6.93	0.764

### 5.2.2.3 The difference in change between the two apps for objective outcomes and the period effects

The results of the secondary endpoint of interest showed that there were no differences between the objective PA outcomes between the two apps assessed (Table 45). However, there was evidence of period effect for daily PA count ( $p=0.010$ ), wear time ( $p= 0.030$ ), SB ( $p=0.006$ ), and moderate PA ( $p= 0.051$ ). Taken together, these findings are consistent with the results displayed in Table 44. There was no difference in the amount the apps changed PA, but there was a difference between the activity change from Period 1 to Period 2. The largest improvement was seen with the 1<sup>st</sup> app assessed.

Table 45: The difference in change between App A and B using objectively measured PA including period effects

	Difference between App A and B			p-value	Difference between Period 1 and 2
	LSM	95% CI lower	upper		p-value
PA count	-14.58	-62.50	33.35	0.547	0.010
MVPA	-0.77	-8.79	7.25	0.849	0.071
Light	-3.84	-13.97	6.29	0.453	0.680
Moderate	0.11	-7.38	7.61	0.976	0.051
Vigorous	-0.93	-2.96	1.11	0.368	0.842
SB	16.60	-16.31	49.51	0.319	0.006
Step count	-362.16	-1406.01	681.69	0.492	0.090
Wear time	12.08	-21.46	45.62	0.476	0.030

**The proportion of participants that increased their time in MVPA by 20% from baseline**

In Period 1, 31.4% (16/51) increased their MVPA by 20% (95% CI= 19.1% to 45.39). In Period 2, 26.8% increased their MVPA by 20% (95% CI= 14.2% to 42.9). The CIs for these results exclude 0 suggesting that these results were significant (Table 46).

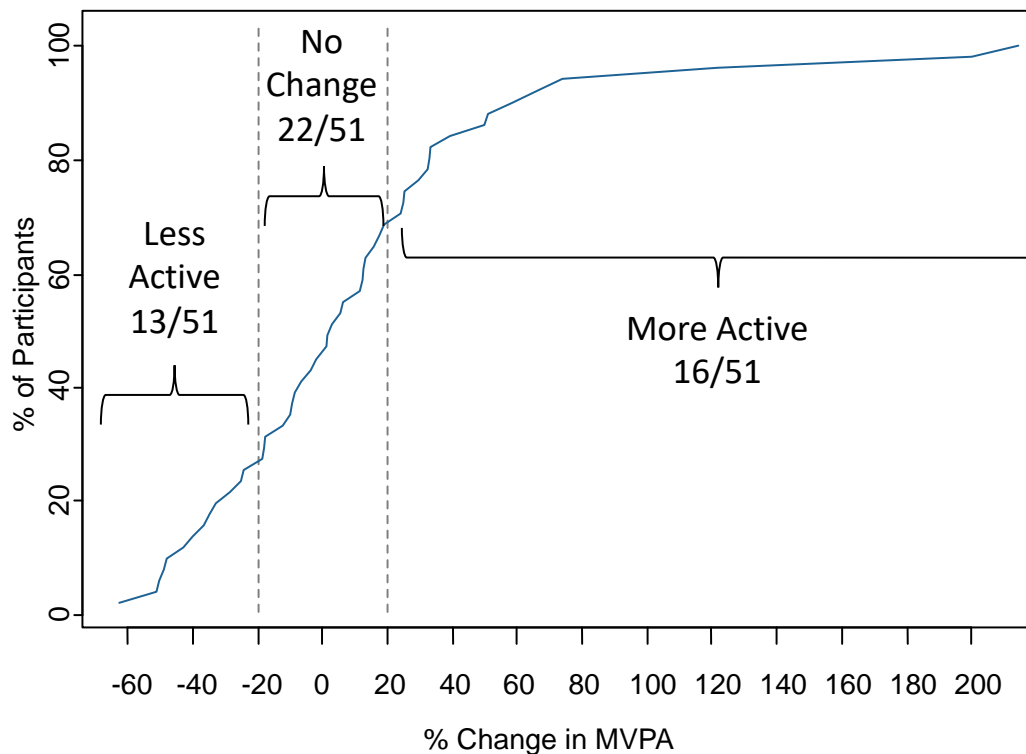
Table 46: the proportion of those participants that increased their MVPA for both periods.

	<b>Proportion of increase in time spent in MVPA by 20% from baseline</b>		
	No increase	20% increase in MVPA	Total available
Period 1	35	16 (31.4%)	51
Period 2	30	11 (26.8%)	41
Total	65	27(29.3%)	92

**Variability in the outcome measure**

Considering the variability in the results of the continuous and categorical outcome, the MVPA outcome was explored further. The distribution of change in MVPA shows the proportion of participants who increased, decreased and did not change their PA (Figure 41).

Figure 41: distribution of change in MVPA showing the proportion of participants who increased, decreased and did not change their PA (Period 1)



***Sensitivity analyses including participants that engaged with the app interventions***

Analysis was conducted on participants that engaged with the app intervention. The median for App A was 21 min of usage and the median for App B was 30 min. Twenty eight participants reported median or above amount of use for App A and 30 participants reported median or more amount of usage for App B. This sensitivity analysis showed that for App A, the results of the primary analysis showed significant difference between PA count, moderate PA, MVPA, and step count from baseline following introduction of the intervention in Period 1 for those who engaged with the App A. In Period 2, SB showed significant increase which could suggest trial fatigue. No significant effects were seen for App B (Table 47 below).

Table 47: Sensitivity analysis showing the primary analysis of the effect of the apps on those that engaged with the interventions for App A and B (in minutes)

	Period 1			p-value	Period 2			p-value
	LSM	95% CI			LSM	95% CI		
		lower	upper			lower	upper	
<b>App A</b>								
Daily PA count	75.47	6.84	144.1	0.030	-42.49	-108.43	23.45	0.200
MVPA	10.8	1.08	20.52	0.030	-0.05	-9.39	9.29	0.990
Light	1.13	-16.22	18.48	0.890	0.59	-16.06	17.24	0.940
Moderate	11.43	1.98	20.87	0.020	0.04	-9.03	9.12	0.990
Vigorous	-0.59	-2.49	1.3	0.520	-0.12	-1.94	1.7	0.890
Step count	1335.76	81.75	2589.77	0.040	-	-	925.5	0.640
SB	-43.17	-99.33	12.99	0.130	83.97	30.03	137.9	0.000
<b>App B</b>								
PA count	-12.7	-89.27	63.87	0.74	48	-31.49	127.5	0.22
MVPA	-1.23	-13.07	10.6	0.83	6.41	-5.89	18.7	0.29
Light	-3.78	-16.03	8.46	0.53	9.74	-2.98	22.46	0.13
Moderate	-3.23	-12.98	6.52	0.5	5.9	-4.23	16.04	0.24
Vigorous	1.87	-2.26	5.99	0.36	0.64	-3.64	4.92	0.76
Step count	-166.48	-	1231.54	0.81	906	-545.88	2357.88	0.21
SB	10.11	-25.2	45.41	0.56	-13.6	-50.27	23.07	0.45

***Sensitivity analyses assessing the relationship between the daily PA count and daily weather***

**Snow**

The sensitivity analysis showed there was no statistical evidence for an interaction between snow and daily PA count by app. This suggests that the PA did not differ between the days with occurrence of snow/no snow by app (75.67, 95% CI (-19.16 to

170.49,  $p=0.118$ ). Secondly, a model without the interaction was run to explore if the snow affected the PA count during the intervention period. The results showed that there was an impact of snow on PA. Specifically, on average, participants did 69.69 more units of PA count (average CPM) when it was not snowing (95% CI: 22.23 to 117.15,  $p=0.004$ ).

### **Rain**

The sensitivity analysis showed there was no statistical evidence for an interaction between rain and PA by app suggesting that PA did not differ between the days with occurrence of rain/no rain by app (LSM=-33.46, 95% CI -99.42 to 32.49,  $p=0.319$ ). Secondly, a model without the interaction was run to explore if rain affected the PA during the intervention period. The results showed no impact of rain on PA count during the intervention period (26.64, 95% CI: -6.12 to 59.41,  $p=0.110$ ).

### **Temperature**

A small effect of temperature in the intervention period was found with each increase in mean daily temperature increasing PA count by 3.28 units (95% CI -0.15 to 6.72,  $p=0.061$ ).

### ***Sensitivity analyses using vector magnitude CPM***

The results of the primary analysis using vector magnitude CPM showed similar results to the ones observed using Axis 1 CPM (daily PA count). The results are presented in Appendix M.

#### **5.2.2.3.1 Self-reported PA, ESE, exercise intentions, PA OE**

The analysis of the self-reported PA outcomes showed significant change (Table 48). Specifically, total time spent in PA, moderate activity, walking, and total PA all increased from baseline to Period 1. Sedentary behaviour decreased. All psychological variables

showed a significant difference. Specifically, exercise intentions and ESE increased. There was a small but significant decline in expected outcomes for PA. Unlike the objective measures, the Period 2 results, with the exception of vigorous PA, were similar in magnitude to Period 1.

### **IPAQ distribution**

Studies using the IPAQ measure have found the results to be skewed and, hence, IPAQ data is often analysed using a log transformation [306, 307]. In this study, the change variables: Vigorous PA, Moderate PA and SB were log-transformed to examine if the results were affected by the skew (See section: Sensitivity analyses conducted on the self-reported outcomes).



## OE distribution

The distribution of variable change in OE was highly skewed (coefficient of skew -2.8, kurtosis 16.5). The inspection of the results showed one extreme case with -33 change from baseline. This outlier was excluded from the analysis and sensitivity analysis was performed including this variable (See section: Sensitivity analyses conducted on the self-reported outcomes).

Table 48: Primary analysis - subjective measure of change in PA: Change in self-reported PA, ESE, exercise intentions, PA OE from post-baseline to Period 1

	Period 1				Period 2			
	LSM	95% CI		p-value	LSM	95% CI		p-value
		lower	upper			lower	upper	
Total time in PA (min/week)	32.52	10.18	54.87	0.005	25.16	2.39	47.93	0.031
Moderate PA (MET-min/week)	113.68	14.97	212.38	0.024	98.49	-2.10	199.07	0.055
Vigorous PA (MET-min/week)	1.85	-201.86	205.56	0.986	37.77	-169.82	245.36	0.719
Walking (MET-min/week)	375.01	106.17	643.85	0.007	366.85	92.88	640.81	0.009
Total PA (MET-min/week)	489.46	118.18	860.75	0.010	504.22	125.86	882.59	0.010
SB (min/week)	-123.23	-196.12	-50.34	0.001	-144.77	-218.48	-71.05	0.000
Expected Outcomes for PA*	-1.70	-2.81	-0.59	0.003	-1.20	-2.33	-0.06	0.039
Exercise intention	5.23	4.86	5.61	<.0001	5.03	4.65	5.406	<.0001
ESE	41.78	38.58	44.98	<.0001	43.20	39.94	46.47	<.0001

Note: \*: 1 outlier excluded

#### 5.2.2.4 The difference in change between the two apps for self-reported outcomes and the period effects

The results of the secondary endpoint of interest shown that there were no differences between the self-reported outcomes between the two apps (Table 49). In addition, there were no period effects suggesting that the order of receiving the interventions did not have an impact on the self-reported outcomes.

Table 49: The difference in change between app A and B in self-reported PA, ESE, exercise intentions, PA OE including period effects

	Difference between App A and B			App A vs App B	Difference between Period 1 and 2
	LSM	95% CI lower	upper	p-value	p-value
Total time in PA (min/week)	-5.72	-37.55	26.12	0.722	0.117
Moderate PA(MET-min/week)	-21.91	-163.63	119.81	0.760	0.318
Vigorous PA(MET-min/week)	121.64	-171.21	414.48	0.412	0.796
Walking (MET-min/week)	-45.24	-433.15	342.67	0.818	0.914
Total PA (MET-min/week)	49.20	-486.53	584.93	0.856	0.966
Sedentary (min/week)	13.36	-91.54	118.26	0.801	0.871
Expected Outcomes for PA*	0.87	-0.61	2.34	0.248	0.590
Exercise intention	0.09	-0.45	0.62	0.752	0.184
ESE	1.13	-3.43	5.70	0.623	0.131

Note: \*: 1 outlier excluded

#### 5.2.2.4.1 Sensitivity analyses conducted on the self-reported outcomes

##### ***The association between the baseline PA and the change in PA in the intervention period***

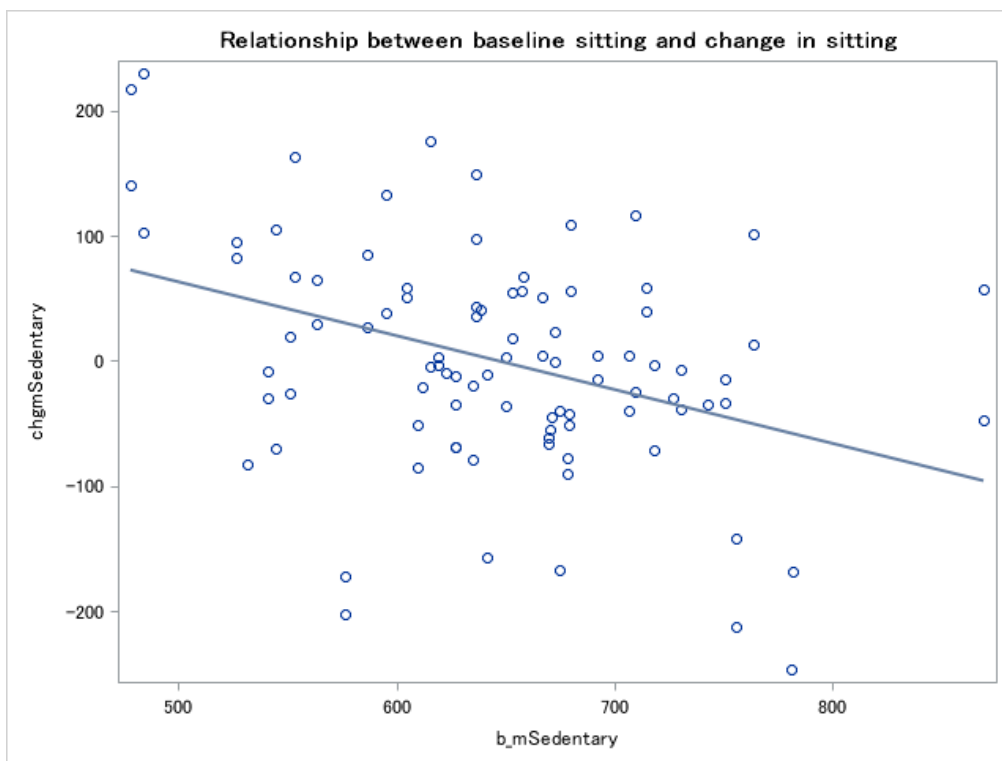
The results of the primary analysis showed that the level of baseline activity affected the change observed in the intervention period. The results of the linear regressions showed that higher baseline activity was associated with smaller increase in PA. Conversely, less baseline activity was associated with larger increase. This result suggests that those participants that were doing least at baseline seem to increase their activity most after the apps were introduced. For example, the results of the SB outcome (Table 50) suggests that for each minute baseline sitting increased, there was an associated reduction in sitting of 0.425 min, i.e., each additional minute of baseline sitting was associated with 25.47 seconds less sitting during the intervention period (60 sec\*0.425).

Table 50: Linear regression of the association between baseline and objectively measured PA and the change during the intervention period

	Coefficient	Standard error	p-value
Daily PA count	-0.36	0.10	0.001
Light	-0.45	0.07	<.0001
Moderate	-0.33	0.09	0.001
Vigorous	-0.20	0.09	0.033
SB	-0.42	0.11	0.000
MVPA	-0.33	0.10	0.001
Step count	-0.33	0.10	0.001

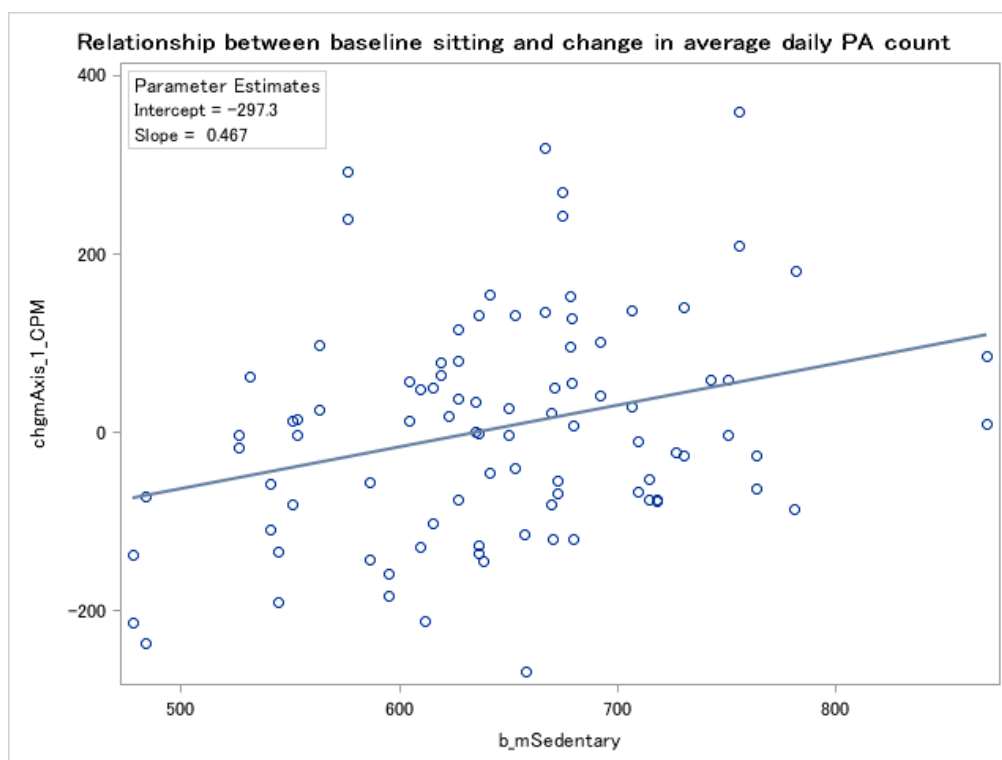
The relationship between the baseline SB and change in SB following the introduction of the interventions can be visualised in Figure 42. The result is showed that more SB at baseline was associated with less sitting during the intervention period.

Figure 42: Scatterplot with a regression line displaying the relationship between baseline SB and change in SB



Linear regression was used to explore the relationship between SB at baseline and PA to explore if those that engaged in more SB at baseline increased their PA during the intervention. The results visualised in Figure 43 and showed that for every additional minute of SB at baseline the PA improved by 0.467 CPM ( $p=0.004$ ). This suggests that the participants who were sitting more at baseline increased their PA during the intervention period to a greater extent than those who were sitting less.

Figure 43: Scatterplot with a regression line displaying the relationship between baseline SB and change in PA



### ***Sensitivity analyses for the self-reported outcomes***

#### **IPAQ**

In the current sample, log transformations were conducted on the IPAQ measure. The results showed similar results to the main analysis (presented in Appendix N).

#### **PA OE**

Sensitivity analysis was also conducted for 1) all records (including the extreme case), 2) with the outlier set to the 2<sup>nd</sup> most extreme value of -16 change in OE. The results were similar to the main analysis which suggests that the outlier had no significant effect on the results (presented in Appendix O).

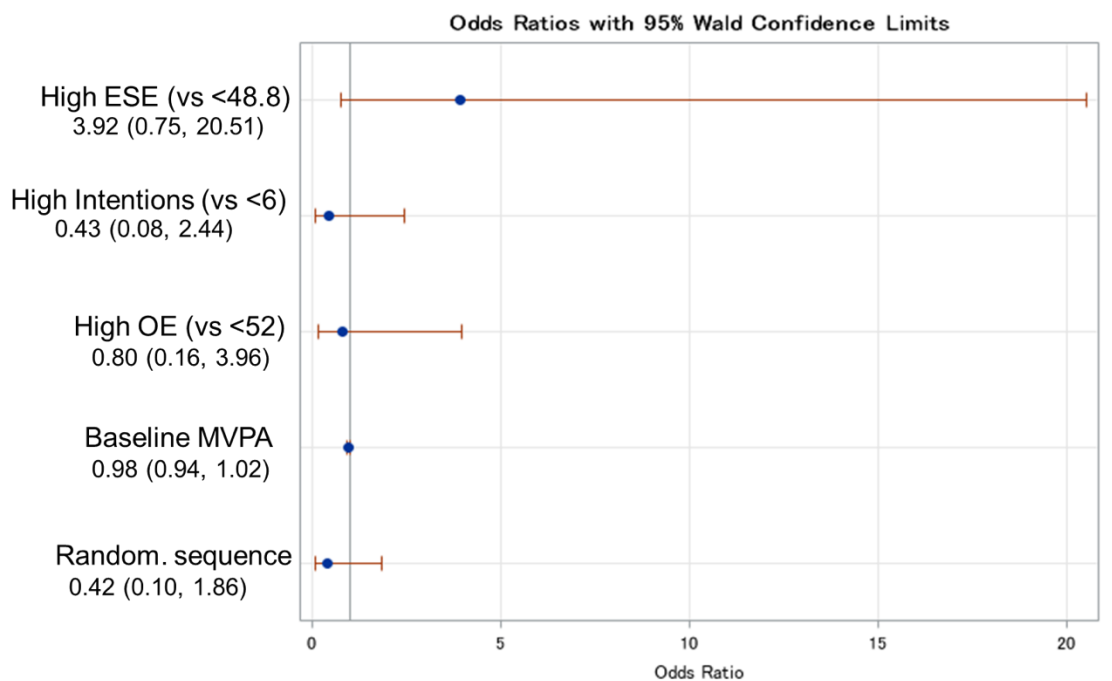
#### **5.2.2.5 Analysis to assess the relationship between the psychological variables and increase in 20% MVPA**

An analysis to assess if the psychological outcomes (ESE, Intentions, and OE) predicted the likelihood of achieving 20% increase in MVPA was conducted. For this analysis, each of the variables was split above and below the median.

A logistic regression including those 3 variables was run, adjusting for the randomisation sequence and baseline MVPA. The results showed no significant predictors of PA. See figure 44.

The 20% improvement in MVPA for those above the median ESE compared to below the median was 3.9 (95%CI 0.75 to 20.1),  $p=0.106$ . This is not strong evidence that ESE is a predictor of 20% improvement in MVPA but it is the only psychological outcome in the expected direction, more related to the outcome than Intentions or OE ( $p=0.34$  and  $0.78$  respectively).

Figure 44: Graph showing the results of the assessment of potential mediating variables on the likelihood of achieving 20% increase in MVPA



Notes: ESE, exercise self-efficacy; OE, outcome expectancy; MVPA, Moderate-vigorous physical activity

## 5.2.2.6 User experience of the apps

### 5.2.2.6.1 Usability of the apps

The average SUS score for the two apps was similar (Table 51), 74 and 75.5 for App A and B, respectively. A score of  $\geq 72.5$  is considered as good usability [165].

Table 51: Descriptive statistics for the SUS score

SUS	N	M	SD	Min	Max	Median
App A	53	74	16.1	37.5	100	75
App B	53	75.5	16.3	35	100	75

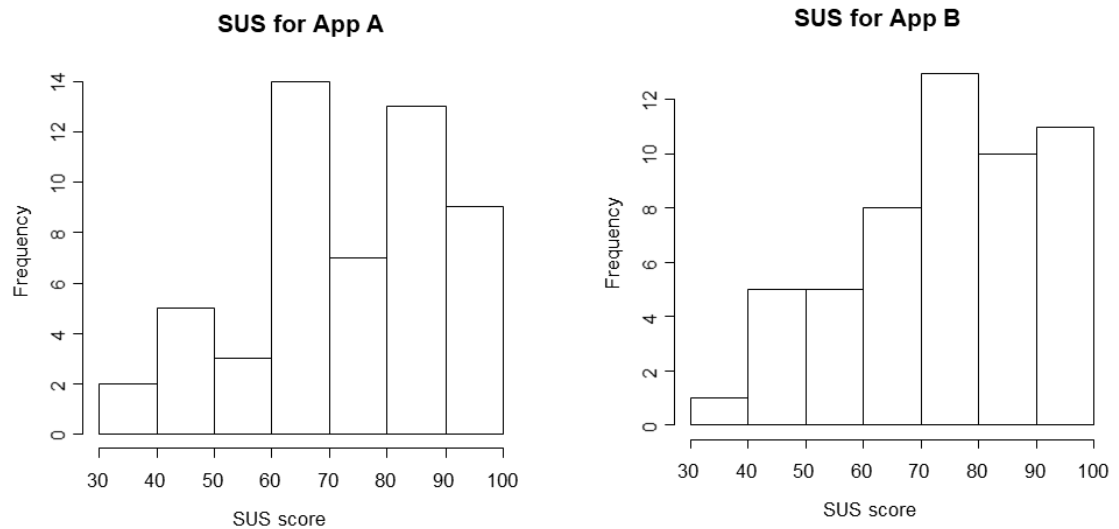
The frequency table of individual scores (Table 52) showed that 49.6% (App A) and 41.5% (App B) of participants indicated a score of 72.5 or lower for each app which suggests that around half of the participants indicated that the apps had good usability.

Table 52: The frequency of SUS scoring for both apps with a score of 72.5 to mark the threshold for good usability

<b>App A</b>	<b>Frequency</b>	<b>Percent</b>	<b>App B</b>	<b>Frequency</b>	<b>Percent</b>
37.5	1	1.89	35	1	1.89
40	2	3.77	47.5	2	3.77
42.5	3	5.66	50	6	11.32
47.5	4	7.55	52.5	7	13.21
50	7	13.21	55	8	15.09
55	8	15.09	57.5	9	16.98
57.5	9	16.98	60	11	20.75
60	10	18.87	62.5	15	28.3
62.5	13	24.53	67.5	16	30.19
65	15	28.3	70	19	35.85
67.5	19	35.85	72.5	22	41.51
70	24	45.28	75	28	52.83
72.5	26	49.06	77.5	31	58.49
75	27	50.94	80	32	60.38
77.5	31	58.49	82.5	34	64.15
82.5	35	66.04	85	35	66.04
85	40	75.47	87.5	40	75.47
87.5	43	81.13	90	42	79.25
90	44	83.02	92.5	44	83.02
92.5	47	88.68	95	48	90.57
95	50	94.34	97.5	51	96.23
97.5	52	98.11	100	53	100
100	53	100			



Figure 45: The distribution of SUS for App A and App B



As SUS variable was normally distributed, the difference in mean usability was assessed using Student t-test. The results showed no difference in mean usability between the apps (-0.55, SD=17.41, 95% CI: -5.50 to 14.54,  $p=0.824$ ).

#### 5.2.2.6.2 Star ratings

The mean star rating for App A was 3.29 (SD 1.12) and 3.4 (SD 0.97) for App B (Table 53).

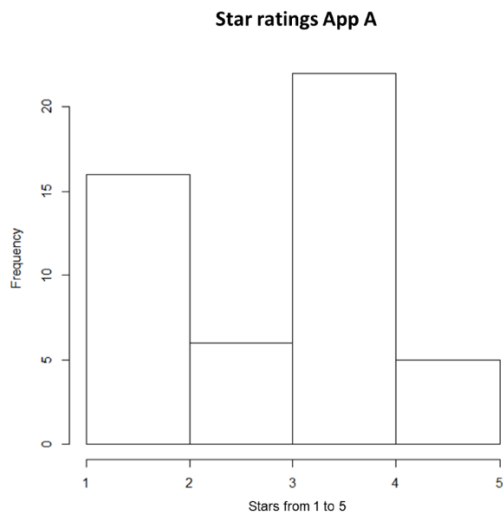
Table 53: Star ratings assigned to each of the apps

Star ratings (1-5)	N	M	SD	Min	25% tile	Median	75% tile	Max
App A	49	3.29	1.12	1	2	4	4	5
App B	47	3.4	0.97	1	3	4	4	5

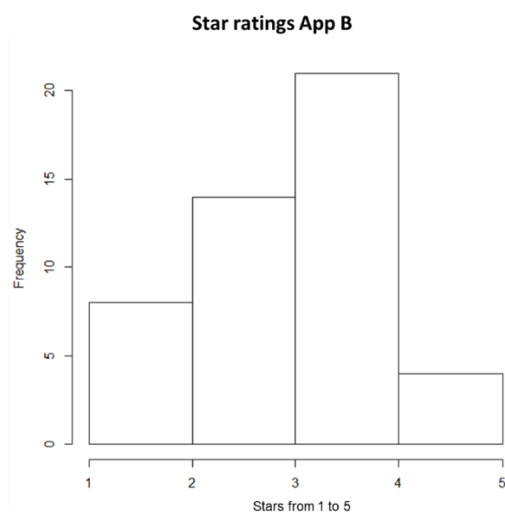
The results of the Student t-test showed no difference in mean star ratings between the apps (-0.07, SD=1.49, 95% CI: 0.53 to 1.23,  $p=0.760$ ).

Figure 46: Distribution of user ratings for App A and B

a) App A



b) App B



### 5.2.2.6.3 Engagement with the apps

#### Experiential items

Table 54 and 55 show the descriptive statistics for the experiential items of the DBCI Engagement scale for both apps. Figures 47 and 48 shows the graph of the experiential items reported by the participants. The results were similar for both apps with 33.23 sum score (SD 8.83) for App A, and 31.85 sum score (SD 9.00) for App B.

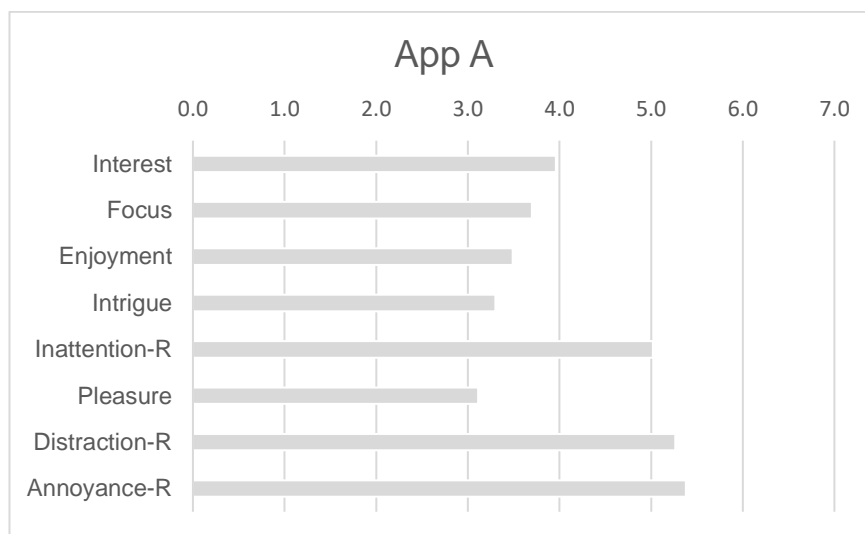
## App A

Table 54: Descriptive statistics for the experiential items of the DBCI Engagement scale (1-7) for App A

<b>Experiential items (N=53)</b>	<b>M</b>	<b>SD</b>
Interest	3.96	1.87
Focus	3.7	1.81
Enjoyment	3.49	1.78
Intrigue	3.3	1.61
Inattention (R)	5.02	1.8
Pleasure	3.11	1.6
Distraction (R)	5.26	1.6
Annoyance (R)	5.38	1.57
Total score	33.23	8.83

Note: (R): items reverse-scored

Figure 47: Visualisation of the experiential items score for App A



Note: R: items reverse-scored

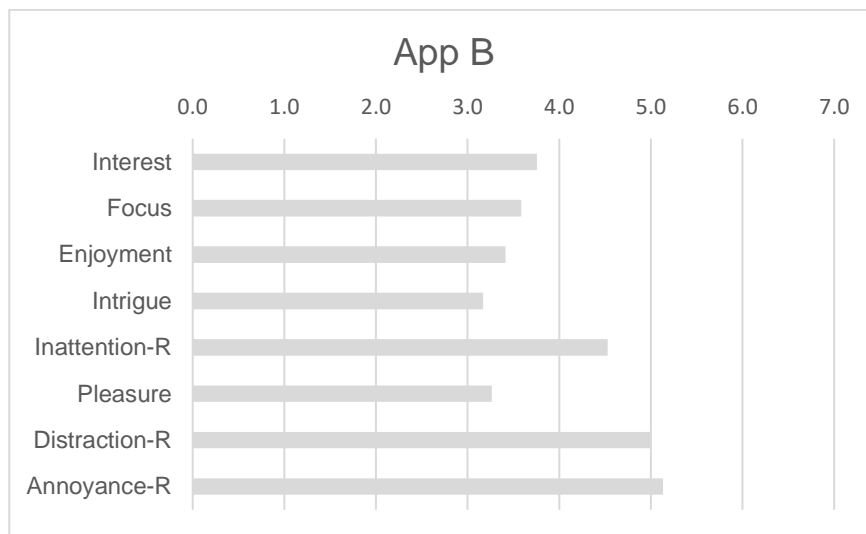
## App B

Table 55: Descriptive statistics for the experiential items of the DBCI Engagement scale (1-7) for App A

<b>Experiential items (N=53)</b>	<b>M</b>	<b>SD</b>
Interest	3.75	1.81
Focus	3.58	1.93
Enjoyment	3.42	1.82
Intrigue	3.17	1.55
Inattention (R)	4.53	1.85
Pleasure	3.26	1.73
Distraction (R)	5.00	1.69
Annoyance (R)	5.13	1.66
Total score	31.85	9.00

Note: (R): items reverse-scored

Figure 48: Visualisation of the experiential items score for App B



Note: R: items reverse-scored

The difference in sum experiential engagement score between the apps was assessed using Student t-test and the results showed no difference between the apps (1.50, SD=10.08, 95% CI: - 1.65 to 9.25,  $p=0.343$ ).

### Amount of use

The median time spent on using the apps was 21 minutes for App A and 30 min for App B. Twenty-five percent of participants spent  $\leq 10$  minutes on both of the apps and 25% spent  $\geq 60$  minutes using the apps (Table 56).

Table 56: Descriptive statistics for the amount of use component of the DBCI Engagement scale

Amount of use (in min)	N	Min	25% tile	Median	75% tile	Max
App A	53	0	10	21	60	280
App B	53	0	10	30	60	360

### Depth of use

Of 8 available components of App A, participants reported using 0 to 8 components, and the median number of components used was 4. For App B, the median was 2 of 5 possible, with the range between 0 to 4 components used (Table 57).

Table 57: Descriptive statistics for the depth of use component of the DBCI Engagement scale

Depth of use	N	Mean	SD	Min	25% tile	Median	75% tile	Max
App A (of 8)	53	3.87	1.85	0	3	4	5	8
App B (of 5)	53	1.72	0.86	0	1	2	2	4

### 5.2.2.7 Ecological Momentary Assessment of mood

There were some technical issues with the PACO software during the duration of the data collection. Some participants reported not receiving the notifications (N= 5) and some self-reported their mood instead (N=4). In addition, a small number of participants received more than 2 notifications. Lastly, for some of the responses, identical data were recorded twice in PACO dataset. In this case, the duplicates were removed (N=105) with 3264 unique responses remaining.

The mean number of records for the baseline period was 11.79 (SD 4.21), for Period 1 was 10.5 (SD 4.12), and for Period 2 was 9.60 (SD 3.59). Table 58 shows the frequency of responses for each period. As mentioned, some of the participants received more than 2 notifications per day. Consequently, for 11.11% participants there were more than 14 responses for the baseline period, 3.45% for Period 1, and 1 participant had more 14 responses for Period 2.

Table 58: The frequency of responses for each period

Baseline					Period 1					Period 2				
N	Frequency	Percent	Cumulative		N	Frequency	Percent	Cumulative		N	Frequency	Percent	Cumulative	
			Frequency	Percent				Frequency	Percent				Frequency	Percent
1	2	3.17	2	3.17	1	1	1.72	1	1.72	1	1	1.92	1	1.92
3	1	1.59	3	4.76	2	2	3.45	3	5.17	2	3	5.77	4	7.69
6	2	3.17	5	7.94	3	1	1.72	4	6.9	3	1	1.92	5	9.62
7	2	3.17	7	11.11	4	3	5.17	7	12.07	4	1	1.92	6	11.54
8	2	3.17	9	14.29	6	2	3.45	9	15.52	5	1	1.92	7	13.46
9	5	7.94	14	22.22	7	1	1.72	10	17.24	6	3	5.77	10	19.23
10	5	7.94	19	30.16	8	4	6.9	14	24.14	8	7	13.46	17	32.69
11	9	14.29	28	44.44	9	4	6.9	18	31.03	9	6	11.54	23	44.23
12	6	9.52	34	53.97	10	11	18.97	29	50	10	3	5.77	26	50
13	15	23.81	49	77.78	11	5	8.62	34	58.62	11	9	17.31	35	67.31
14	6	9.52	55	87.3	12	5	8.62	39	67.24	12	6	11.54	41	78.85
15	1	1.59	56	88.89	13	6	10.34	45	77.59	13	6	11.54	47	90.38
16	2	3.17	58	92.06	14	8	13.79	53	91.38	14	4	7.69	51	98.08
17	2	3.17	60	95.24	15	3	5.17	56	96.55	17	1	1.92	52	100
19	1	1.59	61	96.83	17	1	1.72	57	98.28					
24	1	1.59	62	98.41	25	1	1.72	58	100					
27	1	1.59	63	100										

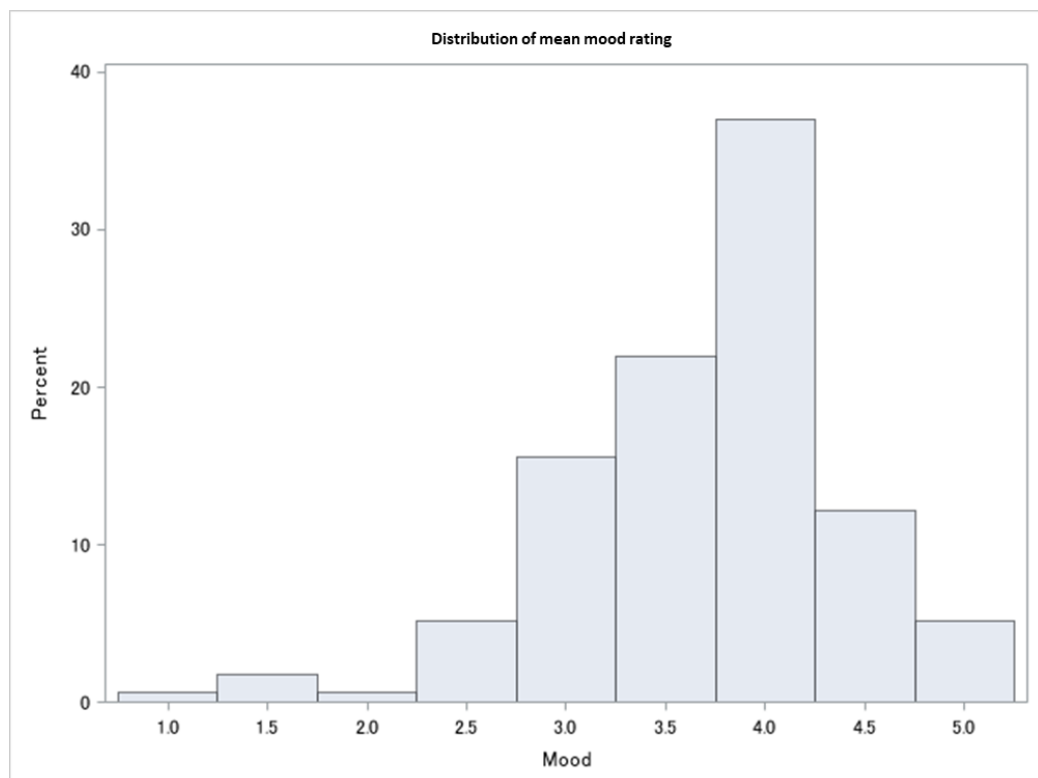
### Average mood in the study

The mean mood was high for all period and for baseline period was 3.59 (SD 0.71), for Period 1 was 3.69 (SD 0.69), and for Period 2 was 3.74 (SD 0.77). See Table 59 for the descriptive statistics of the response number and figure 49 for the visualisation of the distribution of the mean mood responses.

Table 59: Descriptive statistics of the mean mood for each of the study periods.

Study period	N	Mean	SD
Baseline	63	3.59	0.71
Period 1	58	3.69	0.68
Period 2	52	3.74	0.77
Total	173	3.67	0.72

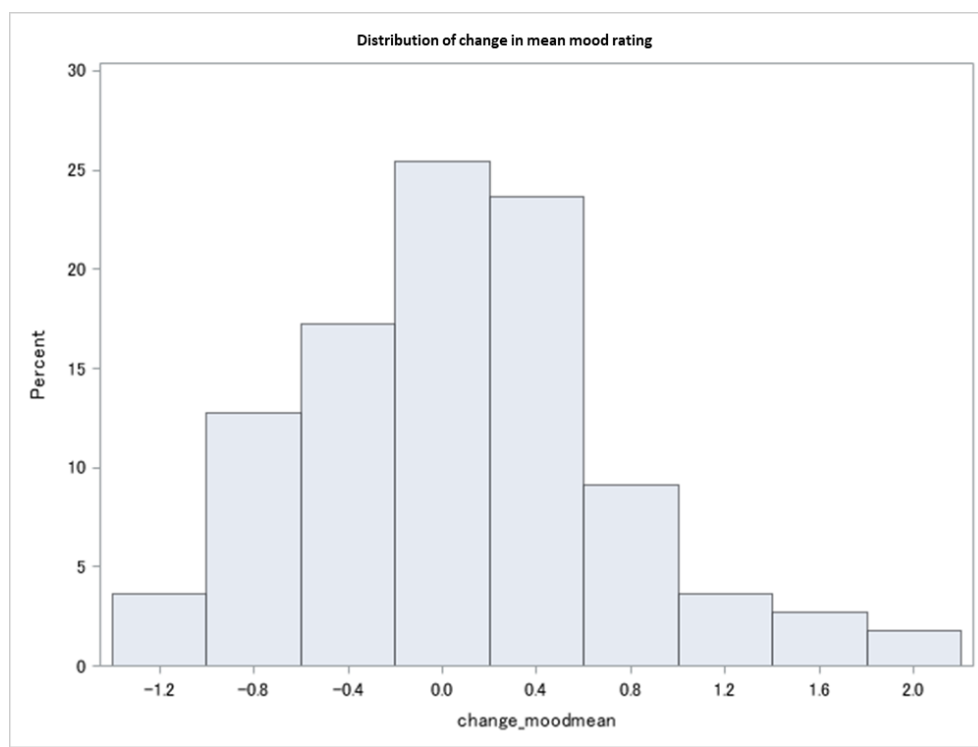
Figure 49: The distribution of the mean mood for all 3 periods



### Difference in mood from baseline to Period 1

There was no change in the central tendency from start of the study to Period 1: the mean was 0.795 (SD=0.661) with both median and mode of 0 suggesting that the average mood of the participants did not change from baseline to Period 1. Wilcoxon signed rank test was used to assess the difference in mood between baseline and Period 1. There was no statistically significant change in mood between baseline and Period 1 ( $Z= 243, p= 0.420$ ). The distribution of the change variable is shown in Figure 50.

Figure 50: The distribution of the change in mood variable



### Difference in PA by mood

The mean mood responses were dichotomised into high (response 4 and 5) and low (response 1, 2, 3) and the descriptive statistics for the objectively measured PA change for mean baseline and Period 1 mood were computed.

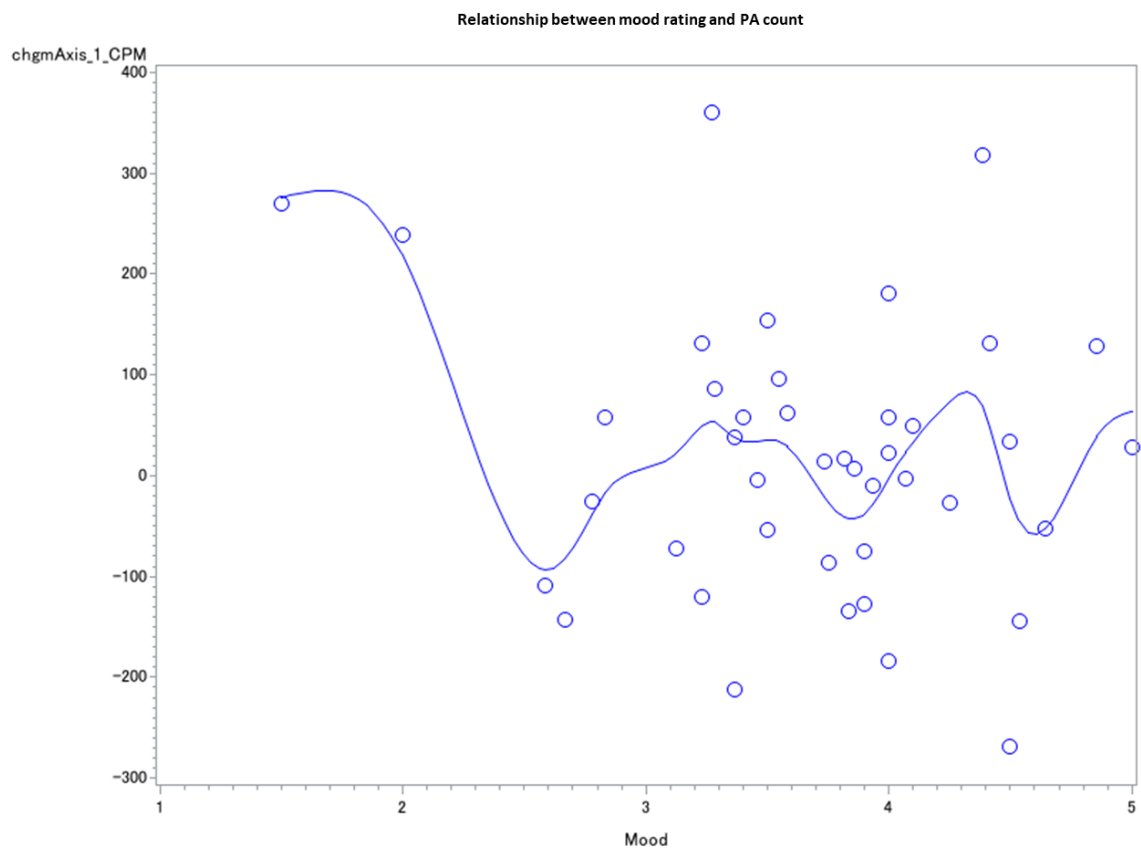
A mixed model for repeated measures was run including fixed effect for app and mood (low/high) with PA count as the dependent variable. The results showed no association



between low/high mood and PA count during the intervention period (-42.45, 95% CI: -102.82 to 17.92,  $p=0.166$ ).

A scatter plot with a smoothing spline was used to explore the functional form of the relationship between PA change and the average mood at the time (average at the time), Figure 51. Two findings can be observed from the plot. First, the relationship between the change in PA and mood is not linearly related. Second, there were 2 extreme cases were participants in low mood highly increased their PA. These cases affected the average relationship between mood and PA change.

Figure 51: Scatterplot with smoothing spline showing the relationship between mood and PA count



## **5.3 Discussion**

### **5.3.1 Principal findings**

The primary objectives of the study were to assess the feasibility and acceptability of a trial to assess popular PA apps available on the market. Both the recruitment and retention rates suggest that such a study would be feasible and acceptable to the participants. The secondary objective assessing the potential impact of the apps on PA showed no effect of the interventions on objectively measured PA using continuous variable. One third of participants increased their time in MVPA by 20%. ESE and intentions increased as expected but PA OE decreased. Baseline PA had an impact on the change in activity observed. Those participants for whom SB was higher at baseline had less SB during the interventions and were engaging in more PA during the intervention suggesting that the study reached those that the interventions was aiming at, i.e., physically inactive participants. There was no difference in outcome between App A and B, and there was no significant difference between the study periods, which supports the use of the crossover design to investigate apps. User experience of the apps varied substantially supporting the notion that one-size-does-not-fit-all. The engagement with the apps was quite low, especially for App B and the interviews with the participants will shed light on the possible reasons for the low engagement rates.

#### **Primary objectives: feasibility and acceptability**

The recruitment process suggested that it is possible to enrol participants to a crossover trial assessing two apps. More than 200 people accessed the screening questionnaire during the recruitment phase. Over 60% of those who were eligible enrolled and were randomised. The completeness rates of the follow-up data were just under 80% for both periods which were high when considering high dropout rates for digital interventions (e.g., [57, 308]).

In research where accelerometry is used, participant adherence rates to accelerometer wear vary depending on the expected length of wear and the population, e.g., [309, 310]. In population level research participant non-adherence to accelerometer protocol (defined as the percentage of those participants that did not meet the criteria for valid data) was between 6 to 32% with a median on 17.6% [311]. In this study, 52 of 66 (just under 80%) participants provided valid data for the accelerometer outcome suggesting high adherence rates and hence feasibility of this data collection method in this population.

The wear time of the accelerometer was higher in Period 2 possibly suggesting that participants did not experience any burden associated with wearing the device. However, PA count and moderate PA decreased and time spent sitting increased in Period 2 which could be an indication of trial fatigue. An alternative explanation could be that participants started reverting to their baseline PA. There is evidence that the effects of behaviour change interventions diminish overtime [312, 313].

### **Secondary objectives: the potential effects of the apps**

#### ***Physical activity outcomes***

The results of the objectively measured PA assessment showed no effects of the app intervention on mean change from baseline to 3 weeks follow-up. However, this study was not powered to identify significant changes. The direction of the changes from baseline to Period 1, with the exception of light activity, were all in the direction of increasing PA. In addition, there was a reduction in SB. These findings suggest that adequately powered trial might show significant impact of the PA apps. In Period 2 the positive results are diminished and the SB was increased compared to baseline.

However, 16/51 (31.4%) participants increased their time in MVPA by 20% from baseline to Period 1. Thirteen of 51 (25.5%) participants showed a decrease in time spent in MVPA by 20% from baseline to Period 1. The confidence intervals for these results exclude zero suggesting a significant difference. This results supports the finding of the research literature that emphasize the need for tailoring of digital health [223]. This could be done through questionnaires exploring the variation across groups in their response to PA intervention. This questionnaire could be then used to tailor the digital technology to users' characteristics and needs.

The variability in the findings between the continuous and categorical outcomes emphasises the inherent issue with using averages to draw conclusions about the efficacy or lack of efficacy in studies assessing PA as an outcome. PA is known to differ between and within individuals and varies throughout the lifespan [314].

Self-reported PA outcome showed different results to the objectively measured PA. Specifically, significant improvements were found for the total time and minutes spent in PA, as well as increase in moderate PA, walking, and a decrease in SB. Unlike the objective measure, the self-reported outcome did not seem to show period effects.

Indeed, self-reported PA has been shown to relate poorly to objectively measured PA and none of these measures is optimal [315, 316]. In addition, the PA measured using the accelerometer device was a daily measure. In comparison, IPAQ asked participants to recall their weekly PA.

Accelerometer device is a state-of-the-art measurement for PA. However, it has its limitations. First, accelerometers might be less efficient in capturing activity that is not step-based, for example cycling [257]. Second, there is evidence that different device

brands are more sensitive to detect various intensities of PA. For example, ActiGraph may not fully capture lower intensity PA and SB [317], and some researchers recommended that this brand is more suitable to be used in research assessing higher PA intensity. Hence, researchers who use one single device should be aware of these limitations of each of the devices. This study used one device, and hence the PA observed, especially of lower intensity, might have been underestimated. Lastly, this study used one of the most commonly used and validated methods for wear time validation [286] and data classification into the intensities of PA [288]. However, the data were processed in ActLife (software developed by the supplier of the ActiGraph). Some researchers have developed their own methods for data processing [285] but this was beyond the scope of this study.

A systematic review of 23 validation studies has shown that the results of the IPAQ short version measure overestimates PA by an average of 84% in comparison to accelerometry-based measurements. The authors argue that the measure should not be used as either a relative or absolute measure of PA. They conclude that it can be used, with caution, in repeated measure studies which was the case in this feasibility trial [318].

This result is a criticism of the IPAQ measure itself but also the nature of self-reporting in PA research. IPAQ is one of the most frequently used questionnaire measuring PA [319, 320]. The overestimation of PA self-report phenomenon could be identified because of the frequency of IPAQ usage and demonstrate the inherent tendency of bias in thinking and recalling of past health behaviours, as described in the next paragraphs.

The effect of over-reporting PA might be due to self-serving bias, i.e., the intention to interpret information to preserve a positive self-image [321]. Public health messaging around obesity and healthy eating has increased. At the same time, rates of self-reported

weight-loss attempts have risen substantially [322]. It has been suggested that this might produce the desire to underreport the calories consumed. In fact, there is evidence that, in national surveys, the calorie intake has been underreported and this tendency has been rising [323]. Similarly, the data produced by IPAQ have been shown to overestimate the PA conducted [306]. As such, it is possible that a similar thinking bias influenced the reported PA in this study.

There are means of addressing this issue of over-reporting using IPAQ with either changing the procedure of the interviews using IPAQ telephone version [324] or adjustments to the data [325]. This was beyond the scope of this thesis but could be considered for a definitive trial. The option to use the IPAQ 31 item long form could also be considered. However, the acceptability of the survey measures, and the survey fatigue in particular needs to be considered when designing the definitive trial. Lastly, more frequent self-reported EMA could be used instead of a retrospective weekly measure which would decrease the recall bias [326, 327].

In addition, when considering those participants who reported using the app for 21 or more minutes for App A, the results showed that daily PA count, moderate PA, MVPA and step count increased following the introduction of the intervention in Period 1. No significant effects were observed for App B. It is possible that the reported engagement with App B was suboptimal to show any impact if there was one. Whereas the median app usage for App A was 21 minutes suggesting that the typical use of this app was three times during the intervention, for App B, 30 minutes was the typical use which would suggest one session per the intervention period. However, the recommended number of session per week is 90 minutes for App B. Hence, as some level of engagement is required for the digital intervention to produce any impact [328], it is

possible that no difference was seen even in those classed as engaged because of the poor usage of this app.

### ***Psychological outcomes***

For psychological mediators of PA, ESE and intentions to do PA increased. These mediators of PA are well established variables that increase engagement in PA. However, small but significant declines in PA OE were reported by the participants from baseline to first intervention (Period 1). The results are not in-line with the premise of the models of health behaviour. Specifically, socio-cognitive models of health behaviour include expected outcomes as an important predictor of behaviour. For example, in Health Action Process Approach, Schwarzer [329] argues that the OE, i.e., the belief that the behaviour will produce a desired outcomes, is a pre-requisite ('preintender' in the model) of intentions to engage in a behaviour. It is possible that participants' baseline expectations were unrealistically optimistic about the outcome of PA prior to starting the intervention. They might have adjusted their expectations following engagement with the PA through the apps. The second explanation for the reduced perceived positive outcome of PA could be related to the overall low engagement with the apps. The cognitive dissonance experienced by the participants who intended to increase their PA but did not engage with the app adjusted their perceived positive outcomes of PA in order to cope with the incongruence between their perceptions of the benefits of PA and the lack of engagement in PA.

The findings of the analysis showed that the baseline PA had an impact on the change observed using the accelerometer. More specifically, those who engaged in less activity during the baseline period increased PA more than those for whom the baseline activity was higher. In addition, more SB at baseline was associated with less sitting during the intervention period. This result was observed across the objectively measured PA.

The results of the analysis that looked into the difference in outcomes between the two apps showed no significant results for either measures. In addition, PA count, moderate PA decreased, and time spent sitting increased in Period 2 which could be an indication of a trial fatigue.

The results comparing baseline and one week post-baseline measures showed that participants reported more vigorous PA, higher intentions, and lower OE at the post baseline assessment. This result suggests that the outcomes in this study were already changing before the intervention was introduced and supports previous findings on the effect of mere introduction of the measurements [330-332]. The wear time of the device was higher in Period 2 possibly suggesting that participants did not experience any burden associated with wearing the device.

This study provided support for the widely observed phenomena of the impact of seasonality and weather on engagement in PA [275, 276]. The analysis of the results taking into account the weather showed that mean temperature had a small but significant positive association between temperature and PA and snow had a negative impact on PA. There was no difference between the apps. This is surprising as App B is more likely to be used outdoors.

The analysis of the predictors of 20% increase in MVPA showed no significant results. However, for those with higher ESE were more likely to have 20% improvement in MVPA. The results was not statistically significant. The mediation analysis for the definitive trial should look at the potential mediators of PA to target the apps to those more responsive to the intervention.



### **App-specific outcomes**

The results of the quantitative user experience outcomes showed that, overall, both apps had good usability as perceived by the participants, with similar scores for both apps (74 for App A and 75.5 for App B). However, the scores varied ranging from a score of 37.5 to 97.5 for App A and 35 to 97.5 for App B which using the descriptive adjectives [165] can be described as from *poor* to *excellent* suggesting variable digital literacy of the sample. The star rating assigned to apps were similar for both apps with App A receiving 3.3 stars on average and App B receiving 3.4 stars. The variability of the findings supports the notion that user experience varies substantially and no-size-fits-all.

When comparing these results to the review and content analysis of the quality of apps (Study 1, Chapter 2), the average of expert led assessment of usability was higher for the sample of apps assessed with a median SUS score of 86.3. The mean star ratings derived from the app stores was also higher with only 25% of ratings below 4 stars.

Although the sample in Study 1 was 65 of the most popular apps, which are not directly comparable to this feasibility trial assessing two apps with multiple users, the discrepancy may suggest that expert led assessment of usability in Study 1 was higher than the rated usability of an average user because of the higher digital literacy of the assessors. Secondly, the star ratings may be inflated for the most popular apps in the commercial app stores.

This study provides some support for the argument that the ratings for the most popular apps in the dominant app stores are inflated. The average rating for both apps in this study was 3.5 stars. In comparison, average ratings are higher for the top-ranked apps in the app store with 75% ratings of 4 stars or above. Relying on app store to provide the most liked apps might be unreliable.

However, the difference between the selection process between apps in app stores and this study is that in app stores, potential users seek out and choose to download an app. In this study, participants were asked to download an app, which they might not have downloaded themselves.

### ***Engagement***

The *experiential items* for the engagement scale showed similar results for both apps with no particular weight for any of the items. However, the *amount of use* showed that the median time spent on using the apps was low, particularly for App B. Each session in App A, as the app name suggests, requires seven minutes to complete but the app does not suggest the frequency of sessions to be completed. App B is more prescribed. It requires three sessions to be completed in each week of the programme, hence App B requires 90 minutes a week to be completed. Yet, the medium reported minutes of the app usage for the duration of each of the intervention (two weeks) showed that for App A 21 minutes was the typical time devoted to the app, which suggests three sessions using app A were completed throughout the intervention. For App B the typical time spent on it was 30 minutes suggesting half a session of usage per week, on average. This seems low.

The *depth of use*, i.e., the distinctive features of the apps used by the participants were, typically, four of eight components and for App A and for App B two of five components. This suggests that participants did not explore all the components available in the apps. The qualitative study shed some light on why the usage rates of both apps were low (Study 4, Chapter 6).

### **Mood assessment using Ecological Momentary Assessment**

The average mood throughout the study was similar with the means in each period of 3.7 and no significant changes were found in the distribution of mood change from baseline to Period 1. This result might suggest that the study procedure or the study itself did not change participants' mood.

In addition, there was no association found between low and high mood and daily PA count. However, the relationship between the change in PA count and mood was not linearly related. There were also two extreme cases where participants with low mood increased PA substantially which make the results difficult to interpret.

Both PA levels and mood vary substantially between- and within- person [333], and the overall effect (e.g. reported using means) is likely to mask this variance [40]. In addition, mood is a state rather than a trait and is characterised by instability and is prone to fluctuation [294-296]. Case-series design offers an alternative that can be used to investigate the variability in PA and mood and assess the relationship between these variables that is more nuanced than comparing the daily averages. The reports in mood could be used to investigate the association between mood and PA. Such studies have been conducted in children [334, 335] and young adults [336]. This was, however, beyond the scope of this project.

### **5.3.2 Implications for the definitive trial learnt from conducting this feasibility study**

Each of the findings will be considered to provide a list of recommendations for the definitive trial:

1. The study was feasible to conduct with representative sample of inactive adults living in urban environment. Although no definitive conclusions can be drawn about the effects of the two apps on PA, the results of this study showed clear differences in self-reported measures: PA, and mediators of PA: self-efficacy and intentions.
2. There was a great variability in the dichotomous outcome (increase in MVPA by 20%). This choice of the threshold was pragmatic for the purpose of exploration in the feasibility trial. Going forward, in the definitive trial, a continuous variable will be used as it is more appropriate (and more powerful) to assess total population impact. However, given the variability that was illustrated using the dichotomous outcome, by looking into sub-groups of participants that increased their PA, there is a scope to explore the characteristics of these participants to predict which groups would most likely be responsive to these digital interventions.
3. The tendency of over-reporting the self-reported PA could be addressed by using more frequent EMA to decrease the recall bias. However, an accelerometer device should be used to measure primary outcome as, despite the limitations discussed here, it is a gold standard in PA research [337, 338].
4. To assess whether the apps can be effective, engagement with these apps is first needed. App A showed some positive results for objectively measured PA within those participants that engaged with this App, App B did not show any positive effect. Engagement (or lack of it) with App B is problematic when considering the definitive trial because low engagement might mean that the app might not be

motivating enough for the participants or it might not be acceptable for other reasons. Independent of the reasons for the low engagement, it is difficult to draw any conclusion on the potential effectiveness of this app and what steps should be taken when designing the definitive trial. In the next Chapter, the qualitative study will give some insights into why the participants typically engaged in only one session throughout the intervention period for App B. Based on the results, improvements to both apps could be made to increase engagement rates with the apps before the definitive trial is conducted. This, however, would require collaboration with the app developers first.

5. In addition, the engagement with the app was self-reported. The definitive trial should use the app usage data to investigate the association between the engagement in PA and the app usage. For this, as above, a collaboration between developers of the apps and the researchers is needed.
6. The mood assessment did not provided any conclusive results but the goal of this study was to assess the feasibility and acceptability of this innovative data collection method. The response rates for EMA of mood were high suggesting that it is a feasible method for data collection. Overall, the measure was highly acceptable to the participants with two exceptions. However, researchers should consider the most appropriate software to deliver EMA. There were several technical issues with the PACO software including some participants not receiving notifications and the responses being recorded twice in the dataset. In addition, there were two participant-reported instances where participants received their notification to respond outside the time set in the PACO settings. In fact, two participants reported receiving the notifications at night. This may deem PACO software to be an unacceptable tool for research.

7. As the usability measure showed variable results, the digital literacy of the participants could be measured at baseline and investigated whether digital literacy has an impact on the engagement with the apps and their effectiveness.
8. As there were some differences found from baseline to post-baseline assessment, it is recommended for the definitive trial to incorporate a no-treatment period to assess whether the impact of the interventions was over and above the change observed by mere participation in the study.
9. There could be period effects as evidenced in the results of this study for the objective measure (accelerometer) but not for the self-reported measures.
10. The study duration of two weeks was sufficient to assess feasibility and acceptability. In the full trial, longer duration will be considered to assess the behavioural effect including PA habit formation. A recent review and meta-analysis looking into PA interventions using wearables and smartphone application ranged from four weeks to 12 month [339].

### **5.3.3 Strengths**

This feasibility crossover trial showed that it is possible to conduct a study assessing the effectiveness of two apps in a population engaging in no or low levels of activity.

The sample in the study was diverse, representing the population of London [340]. There were higher levels of participants' education reported in comparison to the rest of the population in the UK [341]. However, London has one of the highest rates of education in Europe. [342].

This crossover design is a feasible method of assessing the impact of PA apps. The advantage of this design is that it can estimate the overall effect of two apps. In addition, the difference between the two apps can also be assessed.

This study assessed the potential effectiveness of two popular apps as participants were asked to use these in a self-directed manner. This means that the result of the study can approximate the behaviour found in the real world.

This study used a post-baseline to assess the change in PA and PA mediators in the intervention period. As there was some difference between baseline post-baseline, the results might be more conservative but more realistic.

#### **5.3.4 Limitations**

This study has some important limitations. As stated, the purpose of this trial was to determine if it was feasible to randomise patient to the trial, randomise to apps, monitor their PA using an accelerometer and to collect the self-reported measured needed. No statistical inference should be drawn from these findings, and statistical tests should be interpreted as descriptive.

The overall aim of the thesis was to assess the “public health potential of PA apps”. In the process of reducing the design of this trial to a feasible study taking into account the time and the resources available, two apps on the market were selected from the pool of around 48 thousand PA apps (at the time of sample identification for Study 1 in 2016). Although these apps are highly popular and representative of the exercise type targeted in the app store, the result of this study cannot be generalised to other popular PA apps.

The recruitment process included only those potential participants that were classed as engaging in no or low PA. However, self-selection of participants is an issue with intervention studies, and it is possible that the sample of the population that took part in the study was more motivated to engage in PA than those that did not take part.

Due to technical issues, two items of the scale were not recorded correctly. It is unlikely that these items affected the results of the study as they were *missing completely at random*.

The researcher was not able to extract the usage data from the apps and the app developers that were contacted were unable to help with the extraction of the data for the sample of the participants in the study. Hence, the researcher relied on retrospective self-report of engagement with the apps which may have been affected by the recall bias [343].

The limitation of a 20% MVPA was that it was a pragmatic cut-off to be explored in the feasibility trial.

The study duration of two weeks was sufficient to assess feasibility and acceptability, which was the aim of this study. However, two weeks of intervention duration is unlikely to assess real behavioural effects, including habit formation.

Last, the participants in the trial were volunteers, thus introducing a substantial self-selection bias. However, this is an issue with most PA interventions [344].

### **5.3.5 Conclusions**

Overall, this study provided some important findings for a future definitive trial. The recruitment and retention rates found in this feasibility trial suggest that this study is feasible and acceptable to the participants. There are still substantial issues around measuring PA, and the choice of a primary outcome. However, the observed changes suggest that there is the potential to demonstrate some impact of these PA apps on overall PA levels, particularly in participants who were sedentary at baseline.



No impact was observed for objectively measured PA using continuous outcome. There was substantial variability in the PA outcome using the accelerometer device. Mediators of PA showed significant results. The expected outcomes of PA decreased which was an unexpected result. Baseline PA had impact on the change in activity observed. This is an important finding suggesting that those participants who were sitting more at baseline were not just sitting less but they were also engaging in PA during the intervention. There was no difference in outcome between App A and B, and there was no significant difference between the study periods which supports the use of the crossover design to investigate apps.

User experience of the apps varied substantially supporting the notion that *no-size-fits-all* and suggesting different levels of digital literacy of the users. The modification to the future definitive trial include the consideration for the outcome measure, increasing the engagement with the apps, including EMA-type data collection and analysis, and the incorporation of a no- treatment period.

## **CHAPTER 6. Process evaluation of a randomised crossover trial of two popular apps and identification of influences on physical activity: data-prompted interviews**

---

Reflective log

---

### **Changes along the way: Apps and wider context**

As described in the introduction to Part 2 of this thesis, the aim of the qualitative study (as part of the mixed methods study) was to explore the acceptability of the trial procedures and the experience of using the two app interventions trialled.

During the first interview I noticed, and in the following interviews it became increasingly apparent to me, that I needed to look beyond the apps and explore the wider context of activities that the participants engaged in. I realised that I could not explore the factors involved in PA behaviour using the rigid topic guide I developed initially as it addressed the usage of the apps specifically but did not address their fit into daily routines.

Exploring the wider engagement in PA beyond the two apps was important in order to understand how apps facilitated behaviour change and their limitations when considering broader contextual issues of the participants. Hence, I adjusted the topic guide to include a section on a wider context relating to PA (see Appendix P for the interview topic guide). Following this consideration, I decided to focus my analysis on two behaviours: 1) PA engagement in general and 2) PA using the two apps as tested in the feasibility trial.

---

End of the reflective log

---

## 6.1 Chapter overview

In the previous chapter presenting the results of the feasibility crossover trial, a great variability in responses to the interventions was observed. Whilst RCTs can answer the dichotomous question of the efficacy or effectiveness of the intervention, they make the assumption that the intervention will lead to success or failure in a linear way. In reality however, the outcomes are influenced by a multitude of contextual factors. The aim of this study was to explore the “how” and “why” of the observed results of the trial. Process evaluation can help to address these questions. The aim of this study was to first assess the acceptability of the trial procedures, and, second, to identify the influences on PA behaviour and how the app interventions can help or hinder engagement in PA.

Data-prompted interviews were conducted with 36 participants following the completion of the feasibility crossover trial. Purposive sampling was used to select a representative sample of those that engaged and did not engage with the apps. The analysis was conducted until no new data were uncovered to contribute to the analysis, which was reached after analysing 20 interviews. The analysis involved 2 stages: 1) conducting thematic analysis to identify the influences (enablers and barriers) on the behaviour, 2) mapping the influences onto the COM-B and TDF domains.

It was found that the trial design and procedures were acceptable to the participants. A variety of enablers and barriers to PA behaviour were identified and motivation was a major issue around engagement in PA. Moreover, the results helped to understand how the apps could potentially enable or hinder PA behaviour. The two publicly available app interventions can address some of the barriers to PA but may not be sufficient to address the lack of PA at a population level. It is, however, important to target those potential users at the time when they are mostly receptive to PA app interventions.

## **6.2 Background - Methodological considerations**

The background to this qualitative component of *the feasibility and acceptability of a randomised crossover trial assessing two current health apps for increasing PA* describes the methodological considerations when designing this study. This section describes the importance of conducting process evaluation, the choice of frameworks used and the rationale for focusing on two behaviours: experiences of PA behaviour and the app-specific PA behaviour in physically inactive adults residing in urban environments, in addition to acceptability of the trial design and procedures.

### **6.2.1 Why process evaluation?**

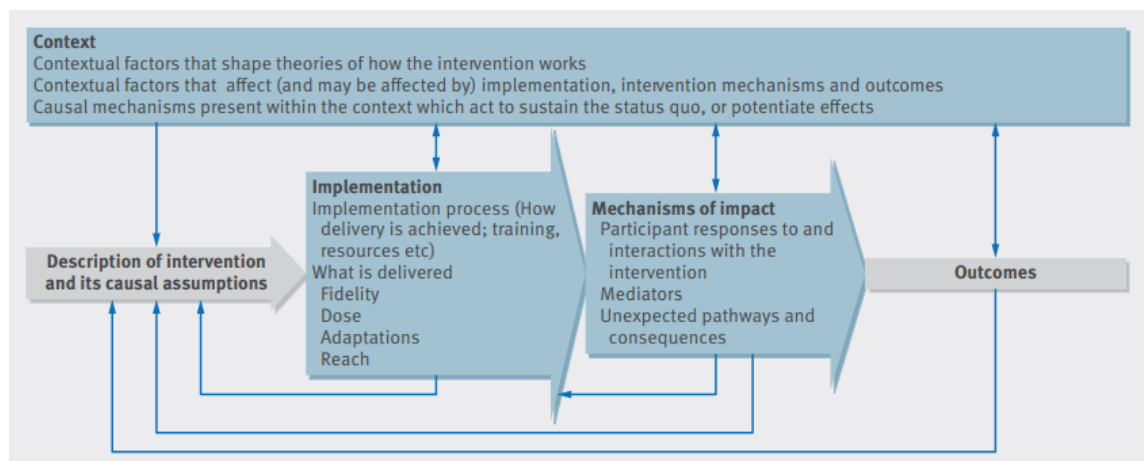
Designing and evaluating complex intervention confers a substantial investment. Whilst outcome evaluation (such as conducted using the “gold standard” RCTs) can answer the dichotomous question of whether the intervention worked. Outcome evaluation, however, cannot inform (i) why and how interventions might work or not, (ii) if the intervention was implemented as intended or whether there were any deviations or adaptations of the interventions, and (iii) what were participants’ responses and engagement with the intervention. These questions can be tackled using process evaluation.

In addition, Moore, Audrey [345] pointed that the effect size, as assessed in RCTs, will not provide information on replicability of the intervention in different contexts, or on the scalability of the intervention. Hence, process evaluation is an important part of any complex intervention evaluation.

The 2008 MRC guidance [107] defines process evaluation as: ‘Study aiming to understand the functioning of an intervention by examining fidelity and quality of implementation, clarifying causal mechanisms and identifying contextual factors

associated with variation in outcomes' (Craig et al. 2008). This comprehensive definition was not accompanied by a guidance of how to plan, conduct, analyse or report process evaluation. In 2015, a new integrative framework was published [345] that builds on the process evaluation themes described in the 2008 MRC guidance (Figure 52 below).

Figure 52: Key functions of process evaluation and relations among them (blue boxes are the key components of a process evaluation. Investigation of the components in shaped by a clear intervention description and informs interpretation of outcomes) (reproduced with permission from [345])



Process evaluation has different functions depending on the stage of intervention evaluation. It also can evaluate different outcomes, as outlined by the definition above.

Specifically to the context of this crossover trial, process evaluation has an important role in understanding the feasibility and acceptability of the intervention and optimising its design and evaluation. The assessment of feasibility using the quantitative methods was described in Section 'Acceptability of the trial design and procedures', Chapter 6).

The guidance on process evaluation suggests that understanding the role of context is part of the process later on when proceeding to the definitive trial. However, as the

overarching aim of this thesis was to investigate the *potential* of PA apps that are available publicly, the contextual factors around PA and app-specific PA behaviour were explored following the feasibility trial.

As I have not developed these interventions (hence I had no control over the content of the interventions and the underlying assumptions about the mechanisms of actions), I decided that, for this study, it was important to explore the apps “in the wild” to explore the influences on PA behaviour in the sample of the population targeted in this study, and how these app intervention can help (or hinder) the engagement in PA.

Hence, the aim of this study were threefold: 1) to shed light on the acceptability of the study design and procedures 2) to explore the contextual influences on PA behaviour (i.e., enablers and barriers), 3) to understand how apps can help (or hinder) engagement in PA.

### **Investigating acceptability**

Acceptability is a complex and multifaceted construct and there is a need to ground the research of acceptability in a theoretical framework [346]. In a recent review of the concept of acceptability, 43 systematic reviews were identified that explored the acceptability of healthcare interventions. However, only 4 provided an explicit definition of acceptability with little theoretical basis or rationale [347]. Sekhon, Cartwright [347] provided a comprehensive and systematically developed definition of acceptability and identified its components (Table 60).

Table 60: The constructs of the theoretical framework of acceptability

<b>Theoretical framework of acceptability</b>	Definition
Affective Attitude	How an individual feels about the intervention
Burden	The perceived amount of effort that is required to participate in the intervention
Perceived Effectiveness	The extent to which the intervention is perceived as likely to achieve its purpose
Ethicality	he extent to which the intervention has good fit with an individual's value system
Intervention Coherence	The extent to which the participant understands the intervention, how it addresses their condition and how it works
Opportunity Costs	The extent to which benefits, profits or values that must be given up to engage in the intervention
Self-efficacy	The participant's confidence that they can perform the behaviour(s) required to participate in the intervention

The framework was used to develop the initial topic guide for the interview schedule. As described in section Summary of Patient and Public Involvement (Chapter 4), two PPI representatives commented on the topic guide. The feedback relating to the topic guide was particularly helpful around the wording of one of the acceptability components *Ethicality*. The phrasing of the question was difficult to understand and was changed following the feedback from the PPI representatives.

### **6.2.2 Interviews with data prompts**

Memory bias is a cognitive bias that alters the content of the reported memory. Data-prompted interviews use personalized multimedia prompts, such as photos, videos, audio recordings, graphs and text. These are used to stimulate discussion in a qualitative interview setting.

The advantages of using data-promoted interviews is their potential to facilitate the retrieval of memory from the past (the retrieval of the contextual memory during the intervention periods). As such, the prompts can stimulate and guide the discussion, complement or contrast participants' responses [348] and hence provide a more holistic and rich data [349].

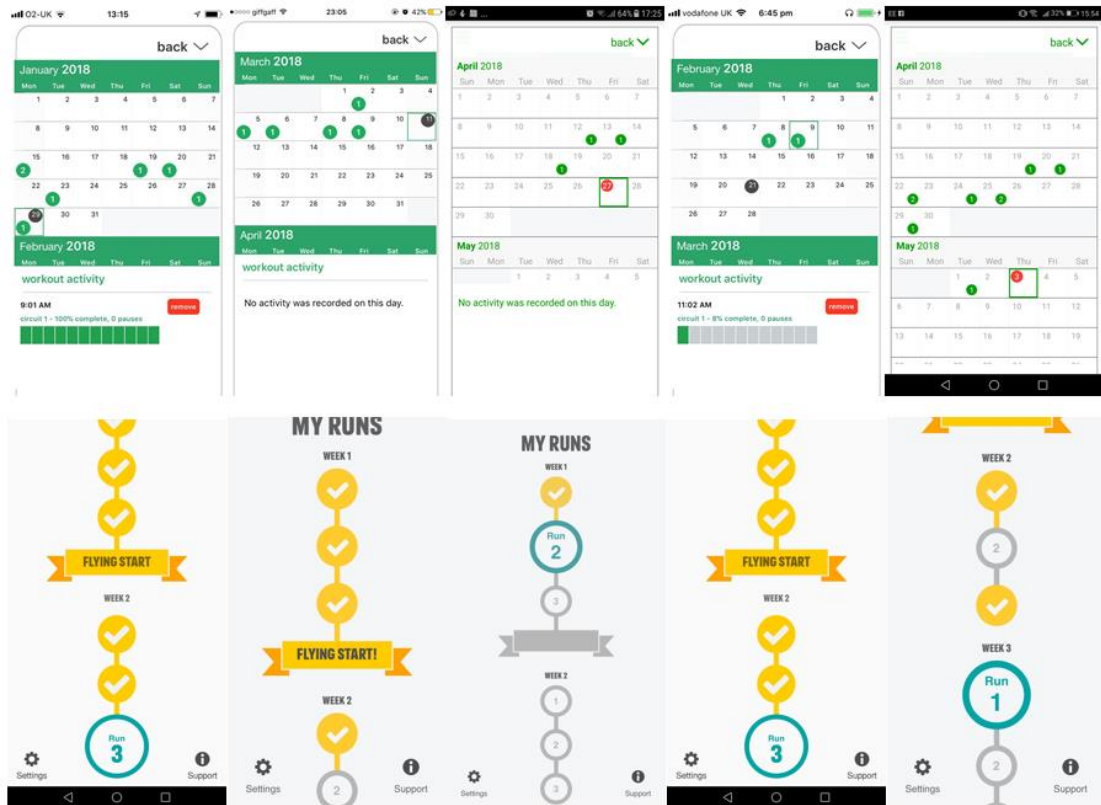
As it was not practical for this study to observe participants in their daily life to understand the usage (or non-usage) of the apps, I used data prompts to facilitate the recall of the experiences of PA and apps.

Data-prompted interviews have been used to gain in-depth understanding of participants' experience, for example, in adherence to medications in Cystic Fibrosis [350], exploring the environmental factors underpinning weight loss maintenance [351], to understanding PA maintenance following cardiac rehabilitation [352], and to generate insight into app usage of smoking cessation app [353].

Three types of data were used to prompt participants' memory and to minimise the social desirability bias that might have occurred in the context of an interview [354]. First, I used participants' survey responses when relevant. An example of the use of the prompt is an interview situation when participant speaks favourably of the app intervention but only rated the app 3 stars in the follow up questionnaire. Enquiring the reasons for this rating prompted the exploration of new areas of the conversation. Secondly, I used the screenshots of activity (Figure 53 showing the sample screenshots) to discuss the amount of usage and, most importantly, the reason for the usage and non-usage. Lastly, the multimedia diaries were used, where available, to evoke contextual memory of the intervention period (Figure 54 shows example diaries sent by participants).



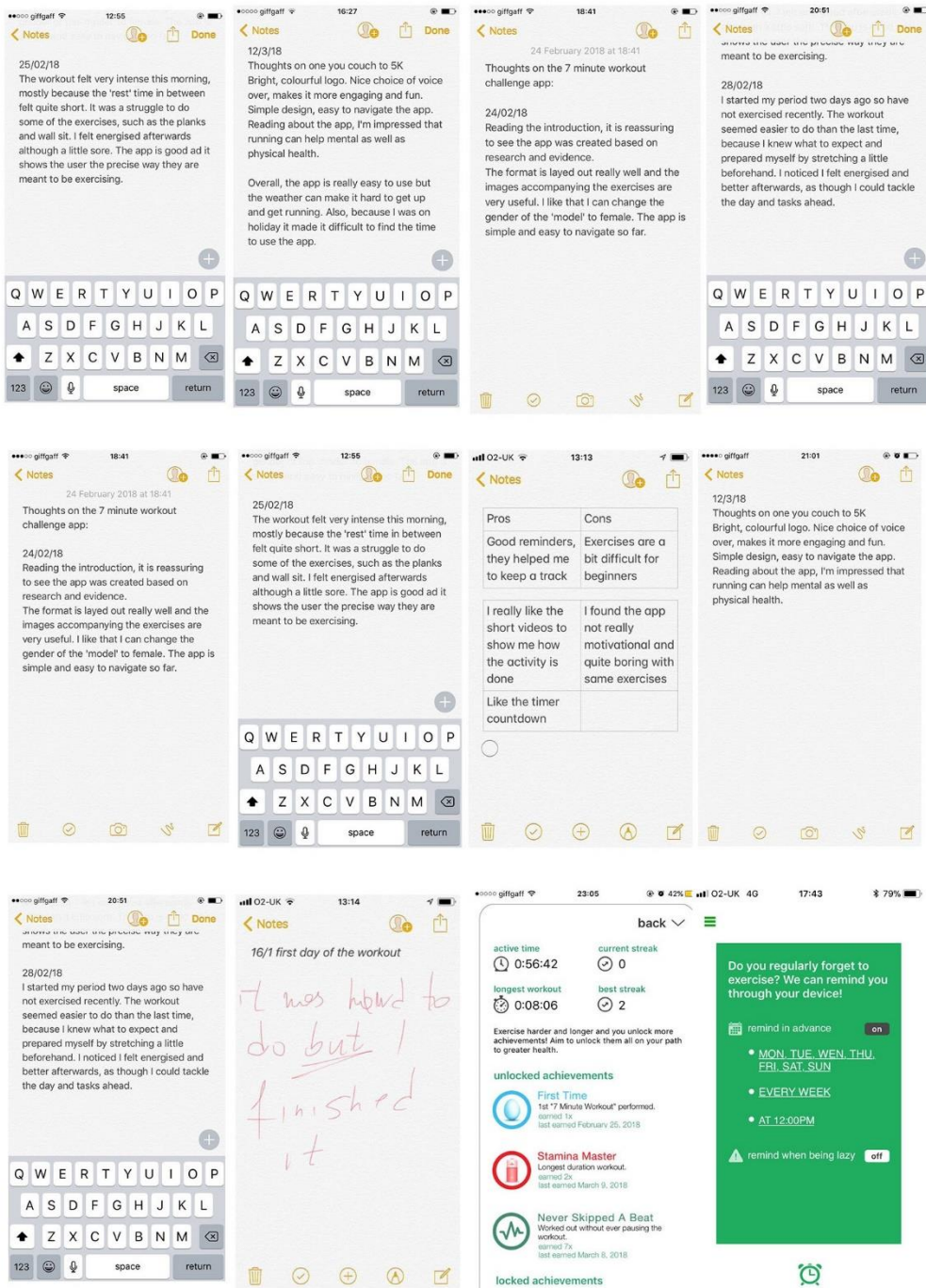
Figure 53: sample screenshots of the usage of the apps used in the interview



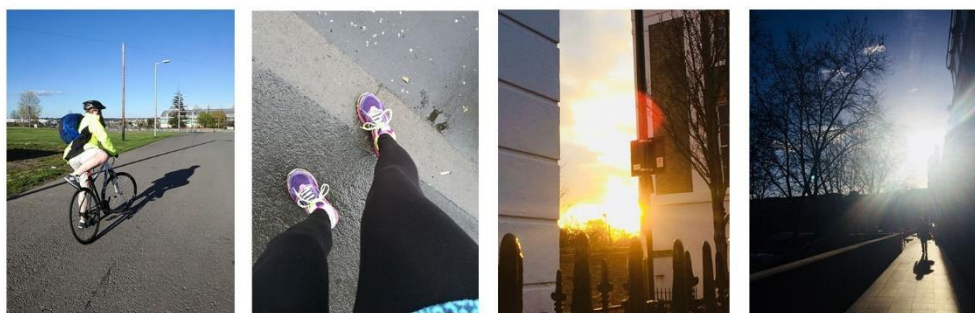
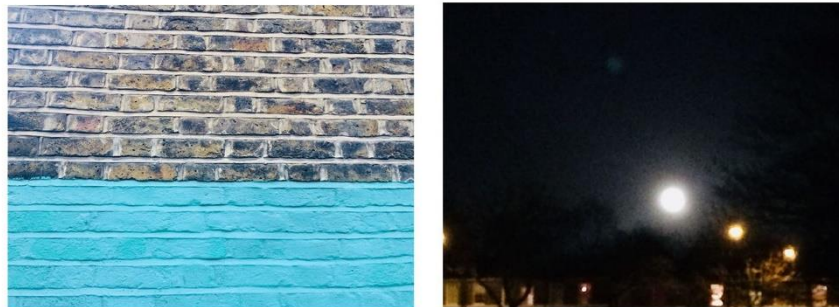
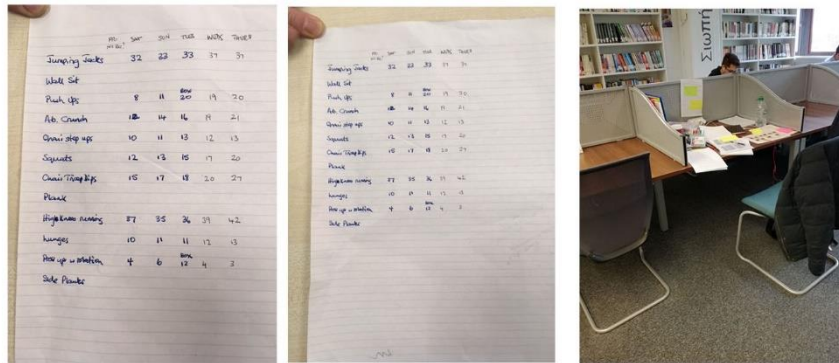
(Source: *7 Minute Workout Challenge* by Fitness Guide Inc. and *One You Couch to 5k* by Public Health England)

Figure 54: sample of the multimedia diaries provided by the participants used in the interview: a) notes and screenshots, b) photos

a) notes and screenshots



b) photos



### **6.2.3 Strategic considerations for data analysis in this study**

The analysis process included two methods. First, the data were analysed following a thematic analysis to identify influences (enablers and barriers) on PA engagement, and how apps can help or hinder PA engagement. The second stage was to map the influences on behaviour onto the TDF and COM-B to help to systematically characterise the areas critical to behaviour change in the context of increasing PA in people who are physically inactive and residing in urban environments.

The rationale for the use of the COM-B and TDF in this thesis was presented in the Introduction Chapter. These tools were used to map the PA determinants/correlates to systematically select the measures for the feasibility trial (Appendix J and K). In the context of this study, the COM-B and TDF were used to categorise and describe in theoretical terms the influences, i.e., enablers and barriers, on PA behaviours, identified in the interview. The use of this framework was particularly important for the next stages of research beyond this thesis that I am hoping to conduct (see section 'Immediate future research direction', Discussion Chapter). The data were described using both COM-B and TDF, with COM-B offering a higher level summary and TDF offering a more granular level of analysis (Labels, definitions and examples are presented in Table 61 for COM-B and Table 62 for TDF).

Table 61: Labels, definitions and examples of COM-B (reproduced with permission from [113])

<b>COM-B model component</b>	<i>Example</i>
Definition	
<b>Physical capability</b> Physical skill, strength or stamina	<i>Having the skill to take a blood sample</i>
<b>Psychological capability</b> Knowledge or psychological skills, strength or stamina to engage in the necessary mental processes	<i>Understanding the impact of CO<sup>2</sup> on the environment</i>
<b>Physical opportunity</b> Opportunity afforded by the environment involving time, resources, locations, cues, physical 'affordance'	<i>Being able to go running because one owns appropriate shoes</i>
<b>Social opportunity</b> Opportunity afforded by interpersonal influences, social cues and cultural norms that influence the way that we think about things, e.g. the words and concepts that make up our language	<i>Being able to smoke in the house of someone who smokes but not in the middle of a boardroom meeting</i>
<b>Reflective motivation</b> Reflective processes involving plans (self-conscious intentions) and evaluations (beliefs about what is good and bad)	<i>Intending to stop smoking</i>
<b>Automatic motivation</b> Automatic processes involving emotional reactions, desires (wants and needs), impulses, inhibitions, drive states and reflex responses	<i>Feeling anticipated pleasure at the prospect of eating a piece of chocolate cake</i>

Table 62: Labels, definitions and examples of TDF domains (reproduced with permission from [113])

<b>Domain</b> Definition	Theoretical constructs represented within each domain
<b>Knowledge</b> An awareness of the existence of something	Knowledge (including knowledge of condition /scientific rationale); procedural knowledge; knowledge of task environment
<b>Skills</b> An ability or proficiency acquired through practice	Skills; skills development; competence; ability; interpersonal skills; practice; skill assessment
<b>Memory, attention and decision Processes</b> The ability to retain information, focus selectively on aspects of the environment and choose between two or more alternatives	Memory; attention; attention control; decision making; cognitive overload / tiredness
<b>Behavioural regulation</b> Anything aimed at managing or changing objectively observed or measured actions	Self-monitoring; breaking habit; action planning
<b>Social/professional role and identity</b> A coherent set of behaviours and displayed personal qualities of an individual in a social or work setting	Professional identity; professional role; social identity; identity; professional boundaries; professional confidence; group identity; leadership; organisational commitment
<b>Beliefs about capabilities</b> Acceptance of the truth, reality, or validity about an ability, talent, or facility that a person can put to constructive use	Self-confidence; perceived competence; self-efficacy; perceived behavioural control; beliefs; self-esteem; empowerment; professional confidence
<b>Optimism</b> The confidence that things will happen for the best or that desired goals will be attained	Optimism; pessimism; unrealistic optimism; identity
<b>Beliefs about consequences</b> Acceptance of the truth, reality, or validity about outcomes of a behaviour in a given situation)	Beliefs; outcome expectancies; characteristics of outcome expectancies; anticipated regret; consequents



<p><b>Intentions</b></p> <p>A conscious decision to perform a behaviour or a resolve to act in a certain way</p>	<p>Stability of intentions; stages of change model; transtheoretical model and stages of change</p>
<p><b>Goals</b></p> <p>Mental representations of outcomes or end states that an individual wants to achieve</p>	<p>Goals (distal / proximal) ; goal priority; goal / target setting; goals (autonomous / controlled); action planning; implementation intention</p>
<p><b>Reinforcement</b></p> <p>Increasing the probability of a response by arranging a dependent relationship, or contingency, between the response and a given stimulus</p>	<p>Rewards (proximal / distal, valued / not valued, probable / improbable); incentives; punishment; consequents; reinforcement; contingencies; sanctions</p>
<p><b>Emotion</b></p> <p>A complex reaction pattern, involving experiential, behavioural, and physiological elements, by which the individual attempts to deal with a personally significant matter or event</p>	<p>Fear; anxiety; affect; stress; depression; positive / negative affect; burn-out</p>
<p><b>Environmental context and resources</b></p> <p>Any circumstance of a person's situation or environment that discourages or encourages the development of skills and abilities, independence, social competence, and adaptive behaviour</p>	<p>Environmental stressors ; resources / material resources ; organisational culture /climate ; salient events / critical incidents; person x environment interaction; barriers and facilitators</p>
<p><b>Social influences</b></p> <p>Those interpersonal processes that can cause individuals to change their thoughts, feelings, or behaviours</p>	<p>Social pressure; social norms; group conformity; social comparisons; group norms; social support; power; intergroup conflict; alienation; group identity; modelling</p>

In short, the use of COM-B and TDF enables to systematically characterise the influences (enablers and barriers) on PA behaviour and to investigate whether these influences have been addressed by the content of the interventions. As such, these frameworks enable to identify the gaps between what *is* targeted by the intervention (e.g., BCTs), and what *should* be targeted by the intervention (influences on PA behaviour).

The COM-B and TDF have been used before to understand the influences on the behaviour using qualitative [355-357] and quantitative methodology [358-360].

### **The rationale for the focus on two behaviours: PA behaviour and app-related behaviour**

The two specific behaviours under investigation were i) PA and ii) the use of two app interventions (trialled in Study 3) to increase PA. The rationale for looking at PA behaviour in general, in addition to using the apps to increase PA, was to understand the factors that influence engagement with PA in people who are physically inactive and residing in urban environments. By investigating both behaviours, there is an opportunity to construct a more holistic picture of enablers and barriers to PA in this population but also to explore the gaps that apps address and do not address, and finally, to provide some recommendations for the developers of digital interventions to consider.

#### **6.2.4 Reflexivity**

I considered the various influences of the researcher upon the research process and the findings [361]. I see myself as someone who acknowledges the difficulty of behaviour change with multiple and repetitive first-hand experience of the intention-behaviour incongruence. I believe that this experience has provided me with the sensitivity and empathy towards the participants who shared their experience in the interview.

I am aware that there might be a stigma connected to health behaviour and, throughout the interview process, I was mindful of the possibility that the interviewees that did not engage with the apps may be concerned about being labelled as *lazy* (or be critical towards themselves) and might want to protect their ego by providing multiple reasons outside their control to accommodate for their lack of engagement with the apps and PA. Hence, I assured each of the participants about the non-judgemental attitude I held. I also mentioned, where appropriate, that I may question their reason for non-engagement



further. I perceived that a good rapport was built with all the participants that took part in the interviews.

### **6.2.5 Aim and objectives**

The aim of this study was to explore the acceptability of the trial, and to understand the issues around PA engagement in physically inactive adults living in urban environments and how apps can influence the engagement with PA (i.e., to identify their potential as behaviour change tools). The specific objectives of the study were:

- To identify acceptability issues with the trial procedures
- To describe the influences of engagement in PA
- To identify the enablers and barriers to engagement in PA
- To identify how these two apps can help or hinder engagement in PA

## **6.3 Methods**

### **6.3.1 Design**

Qualitative data-prompted semi-structured interviews. Data prompts were used to facilitate the recall of the experiences of PA using apps. The interviews allowed for in-depth understanding of acceptability of the study procedures and the interventions, and the exploration of the two behaviours: PA generally, and PA using the two apps trialled.

### **6.3.2 Sample**

Purposive sampling was used to select potential interviewees. Participants that showed frequent usage of the apps, as well as those who used the app infrequently (as assessed

on the DBCI Engagement Scale [272] by the *amount of use* - participants' report on the time spent on using both apps) were interviewed to obtain a diversity of experiences. In addition, a variety of demographic characteristics was considered to capture the views of a range of participants as can influence both PA [263] and engagement in digital interventions [362, 363].

### **6.3.3 Recruitment procedure**

Participants who took part in the crossover trial and consented to being interviewed were included in this study. The recruitment was intended to continue until the researcher gathered a varied data from participants of different socio-demographic characteristics as well as different levels of app engagement.

### **6.3.4 Data collection and procedures**

#### **Source of data prompts**

During the interview, the data that were used to prompt the participants' memory were: 1) their survey responses, 2) screenshot of the activity recorded by the apps, 3) multimedia diaries

In relation to the 2<sup>nd</sup> source, *screenshots* of the app feature that showed the completed PA sessions (*Activity calendar* for App A and *My runs* for App B) were used to prompt the recall of the experiences during the interviews. In relation to the 3<sup>rd</sup> source, *multimedia diaries*, participants were asked to record any reflections of their experiences of the apps using various media of their choice: text, photography, audio and video recordings during the intervention phase of the feasibility trial. This was not a compulsory element of data collection. The data prompts were printed and brought into the interview session and participants were asked to refer to their prompts whenever relevant. In addition, participants were asked to open each app at the point of discussion to trigger their memory.

### **6.3.5 Procedures**

The topic guide was developed with the aid of the theoretical framework of acceptability of healthcare interventions. [347] (Appendix P presents the interview topic guide used).

All interviews were conducted within the Department of Primary Care and Population Health or at the main University College London campus in a pre-booked room. Each participant was reminded of the reasons they had invited for the interview. They were reassured that the recordings and the transcripts of the interview would be anonymized and that no personally identifiable information was linked to the data they provide. They were reminded about the importance of their views and opinions, and were assured that there were no right or wrong answers. The request for permission to record the interviews was repeated before starting each interview.

After each interview, I made notes about my perception about any non-verbal communication and any notable events to understand the context of the interview. These field notes formed part of the dataset.

### **6.3.6 Data analysis**

The interviews were digitally recorded and transcribed verbatim by a professional transcription company. Each transcript was checked against the recordings for accuracy and any identifiable information was removed. Subsequently, NVivo 12, a software package that aids with the organisation of the data, was used for the analysis.

Different types of analyses were used to understand the acceptability of the trial procedures (first objective) and the influences of PA and the use of apps (objectives 2-4).

The analysis of the trial procedures (first objective of this study) focused on identifying any issues with the trial procedures, including the use of two apps and the data collection methods (accelerometer wear, EMA collection, questionnaires). As such, the focus of the analysis was on any negative feedback in order to minimise the issues when the definitive trial is conducted.

As described in the Background of this chapter concerned with methodological consideration, the analysis of the issues with PA engagement and the apps (the subsequent three objectives) followed two steps: 1) initial analysis was conducted using a thematic analysis [364] 2) followed by a mapping exercise using the TDF [115] and COM-B [113, 114].

### **Stage 1: thematic analysis**

The process of thematic analysis followed the guidance outlining the six stages of conducting the thematic analysis [364].

First, I familiarised myself with the data through repeated reading and searching for patterns and meaning. For this analysis, the emphasis was on the influences of PA and whether the apps interventions were seen as facilitating or hindering PA behaviour. Secondly, I generated the initial codes. Third, I searched and identified themes. Fourth, I actively reviewed the themes whilst looking for disconfirming data. This was a continuous process. In addition, specifically to this study, I characterised each theme as a barrier, enabler, or both (if the theme was a combination of a barriers and enabler). I then reviewed and re-defined the descriptive labels for the themes (step five). Lastly, I produced the report summarising the themes and illustrating these with quotes (step six).

Participants' responses were coded only when they related to what they had done with the app and not about what they could do. Data clinic was held in the initial stages of the analysis to explore two transcripts and to discuss the emerging themes. Next, following the initial data analysis, there were two rounds of feedback where the themes with the quotes were presented to two experienced researchers (EM and FH) to obtain agreement and other interpretations of the data. The analysis was refined until agreement was reached.

### **Stage 2: Mapping of the themes onto the COM-B and TDF**

In stage 2 I deductively coded the themes, representing enablers and barriers or both, into the TDF and COM-B domains they were judged to best represent. It was possible for a data point to be allocated to more than one domain where appropriate. The process was examined against the quotes of the participants to increase the congruency of the data with the TDF and COM-B. The initial coding of the data on to the TDF and COM-B was discussed and refined, and checked again by an experienced researcher (GF).

To facilitate the presentation of the data, the summary of the themes with the mapping of the theoretical frameworks was summarised including the representative quotes.

## **6.4 Results**

Of 66 participants that took part in the trial, two did not agree to take part in the interviews. Interviews were conducted with 36 participants. However, the analysis was conducted concurrently until no new data were found to contribute to the analysis. Consequently 20 interviews were analysed. The participants represented a diverse range of demographic characteristics and level of engagement with the apps as well as satisfaction with the apps (Table 63).

Table 63: Profile characteristics of participants that took part in the interview (N = 20)

		N (%)	
Age (years), n (%)	18-24	2 (10)	
	25-34	11 (55)	
	35-44	3 (15)	
	45-54	3 (15)	
	55-64	1 (5)	
Gender, n (%)	Female	10 (50)	
Ethnicity, n (%)	White	15 (75)	
	Asian	3 (15)	
	Black	1 (5)	
	Mixed	1 (5)	
Amount of use (min), n (%)*:	Minutes	N (%)	
	App A (7 min = 1 session)	0-10	4 (20)
		11-30	4 (20)
		31-50	5 (25)
		51-70	2 (10)
		71- 90	2 (10)
		91	2 (10)
	App B (30 min = 1 session)	0-30	12 (60)
		31-60	3 (15)
		61-90	1 (5)
		91-120	3 (15)
		121	1 (5)

*\*as measured using the DBCI Engagement Scale, participants' report of the time spent on using each app*

The following section first reports on the acceptability of the trial, and subsequently presents the themes relating to the influences on PA engagement and how app

interventions helped or hinder PA engagement. A summary table of the themes with the classification using the TDF/COM-B is also presented.

### 6.4.1 Acceptability of the trial design and procedures

It was important that the feedback from the participants relating to the design of the trial, and its procedures (including the length of the study, the schedule, and the data collection methods) was recorded and analysed so that these insights could highlight any areas for development for the future definitive trial. The table below (Table 64) provides a summary of the acceptability of the trial design and procedures to consider when proceeding to the definitive trial.

Table 64: Summary of the acceptability of the trial design and procedures with example quotes

Theme	Sub-theme	Example quote(s) [Study ID, age, gender]
Trial design	“Trying” two apps perceived as beneficial	<i>I think, I think it's a good number, two apps. Because it for example, when I first started, I remember to feel, oh, excited because there were two apps that I didn't know about. I was excited about using it. So, I think two apps is a good number. Maybe three (...) [006-CAT, 26, F]</i>
	Complexity of the design	<i>The third week I thought it was just a case of wearing the recorder again, but I should really have kept doing the exercises. [033-STE, 64, M]</i>
Accelerometer	Ease of use	<i>I never really forgot to put it on. But I was, I tend to forget to take it off at night, because, for example, I would be already in my like the, the sleeping kind of clothes and still doing something at the computer for another three hours or two hours. And then I'll just move to bed and finally I would just fall asleep. Not remembering, like I didn't get to remember that I had this on me, like it was</i>

		<i>never – like I could never really feel that I have it. [028-ADA, 30, F]</i>
	Discomfort around wearing the device	<i>I'm never wearing one of those stupid things again. Honestly, it sticks out. I had to change some clothes I was wearing because I don't have belts... People think you've got something stuck on you. You can see it. It's like walking around with a flipping Tamagotchi stuck to you. [022-MIC, 36, M]</i>
	Unstable position	<i>Yes, that's when you'd notice, yes, or at the end of the day, I'd be like, "Oh I didn't notice that it had moved up." And I don't know if that's because I hadn't put it on properly to begin with or I suppose it just moves up if you're sitting down. And then you have to remember to adjust it, I don't know, when you get up again. [034-SAR, 34, F]</i>
	Suspicious object	<i>You do worry sometimes that somebody's got a, you know, you don't know – a member of the public, it could be a tourist, especially in London, they see a big red thing there around your waist. [045-LIA, 26, M]</i>
	Awareness of monitoring (Hawthorne effect)	<i>I didn't, I don't believe that it changed my behaviour because of it, but I was really aware of it. And being like, "Oh I'm wearing heels, so I'm not going to walk to the tube. [023-ADR, 32, F]</i>
EMA of mood using an app	Nonresponse	<i>I didn't mind that at all. It just popped things up now and then. I don't tend to keep my phone on me during the day so I tend to miss them while I'm at work but I didn't mind doing them otherwise. [022-MIC, 36, M]</i>
	Awareness of the mood in the moment	<i>But yes, I kind of liked it, because you know it stops you in the middle of your day and say, Oh, "So how are you feeling?" And then you know, you're kind of, "How am I feeling?" [064-HAY, 33, F]</i>
	The type of prompting	<i>The barking, annoyed, violent – so I normally have my phone on, on vibrate or silent. But it vibrates my watch when the, when the dog wanted attention. [052-WAR, 44, M]</i>
	Technical issues with the app software	<i>Yes, so I just – the email thing was like, I don't understand why. I can understand why it might be my email, but not my email's password. And yes, just like the barking made me – like I'd be sitting at my desk and it would be going (makes a barking sound) and I'd be like, "Oh." Not even like 'woof'. It was like (loud barking noise). And yes, and it</i>



		<i>just would do it in the middle of the night. [023-ADR, 32, F]</i>
Follow up questionnaires	Awareness of sedentary behaviour	<i>Yes, yes, I've done or how – I mean it was crazy for me to just realise how much I sit. There is that question, how many hours a day you sit. And really, like being super honest, fourteen – that was really crazy, you know, it was just crazy to realise that that is the reality. [028-ADA, 30, F]</i>

#### 6.4.1.1 Trial design

As I was unable to find any previous studies using this crossover design requiring participants to use two apps, I was interested in participants' experience of being asked to try two apps in a sequential order for 2 weeks each.

#### “Trying” two apps perceived as beneficial

Overall, participants reported this as a positive experience as they were able to compare and potentially select the PA they preferred:

*I think I like the fact that there was more than 1 app. Because there's something you can compare it to. I think if it was just one app I guess I wouldn't have minded. I like use a lot of apps anyway. So, I guess I can compare it, like in my head, to like some degree. [010-LUC, 19, F]*

And:

*I think, I think it's a good number, two apps. Because it for example, when I first started, I remember to feel, oh, excited because there were two apps that I didn't know about. I was excited about using it. So, I think two apps is a good number. Maybe three even though. [006-CAT, 26, F]*

## **Complexity of the design**

However, a complex design (1 week baseline, 2 weeks each of the apps assessment, with 3 weeks of accelerometer wear) might create some confusion about the procedures for participants. For example, there was one instance when the participant did not engage with the app during the second week of interventions phase:

*The third week I thought it was just a case of wearing the recorder again, but I should really have kept doing the exercises. [033-STE, 64, M]*

### **6.4.1.2 Accelerometer**

Most of the participants who wore the accelerometer reported they had little issues with it. Some of them mentioned forgetting to wear it at times. However, placing the device by the bedside table helped them to remember. Some participants found the schedule and the reminders to wear accelerometer also helpful.

## **Ease of use**

The below extract illustrates the ease of use of the device. Often participants reported forgetting about it within a few minutes of placing it on their hip. For example:

*I never really forgot to put it on. But I was, I tend to forget to take it off at night, because, for example, I would be already in my like the, the sleeping kind of clothes and still doing something at the computer for another three hours or two hours. And then I'll just move to bed and finally I would just fall asleep. Not remembering, like I didn't get to remember that I had this on me, like it was never – like I could never really feel that I have it. [028-ADA, 30, F]*

There were two cases where the participants reported strongly disliking wearing the device although they still adhered to the wear protocol:

*Yes, I hated it at the end. Especially if you drink a lot of Coke and water you always want to go to the loo and you have this thing pushing you there. So I found that annoying. But as you can see, I was also removing it so I was trying to wait for the very morning. Else, if I had shower maybe the morning and then wear it until bed. [042-MAR, 25, F]*

In addition, the second participant mentioned discomfort and self-consciousness around wearing the device.

*...but I'm never wearing one of those stupid things again. Honestly, it sticks out. I had to change some clothes I was wearing because I don't have belts... People think you've got something stuck on you. You can see it. It's like walking around with a flipping Tamagotchi stuck to you. [022-MIC, 36, M]*

### **Unstable position**

Another problem with the device (that might potentially affect the data gathering) was the issue with the device moving, and especially, rising up. For example:

*Yes, that's when you'd notice, yes, or at the end of the day, I'd be like, "Oh I didn't notice that it had moved up." And I don't know if that's because I hadn't put it on properly to begin with or I suppose it just moves up if you're sitting down. And then you have to remember to adjust it, I don't know, when you get up again. [034-SAR, 34, F]*

### **Suspicious object**

Two participants were worried about the device being perceived as suspicious. This might be especially relevant in the context of a city when considering the heightened level of alert around public transport:

*...you do worry sometimes that somebody's got a, you know, you don't know – a member of the public, it could be a tourist, especially in London, they see a big red thing there around your waist. [045-LIA, 26, M]*

### **Awareness of monitoring (Hawthorne effect)**

A concern around the assessment using accelerometers that was explored in this study was behaviour change because of the sheer awareness of being monitored. Only a few participants reported being more aware, especially at the beginning of the trial:

*I think it made me a bit more aware of what I was doing. I mean sometimes it was a bit like, you know, something else to think about, if I. I mean it was not, not a big deal. It became part of a routine and a pattern. So, it's like wake up, take a shower and then put it on. And then just, I mean you kind of forget that it's there, really. Yes. [064-HAY, 33, F]*

Another example of the awareness of the device and the potential altered behaviour:

*I didn't, I don't believe that it changed my behaviour because of it, but I was really aware of it. And being like, "Oh I'm wearing heels, so I'm not going to walk to the tube. [023-ADR, 32, F]*

#### **6.4.1.3 Ecological Momentary Assessment of mood using an app**

Most of the participants did not have any issues with the EMA assessment of mood. However, some important points were identified, such as the potential of EMA as an

intervention, the preference for the prompts and the technical issues with the PACO software specifically.

### **Nonresponse**

The nature of the prompts (at random times) meant that some of the prompts were not answered, especially during working hours:

*I didn't mind that at all. It just popped things up now and then. I don't tend to keep my phone on me during the day so I tend to miss them while I'm at work but I didn't mind doing them otherwise. [022-MIC, 36, M]*

### **Awareness of the mood in the moment**

There was evidence that the assessment using EMA made the participants more aware of their mood and to reflect on how they felt. For example:

*But yes, I kind of liked it, because you know it stops you in the middle of your day and say, Oh, "So how are you feeling?" And then you know, you're kind of, "How am I feeling?" It's like alright. So, how do you sort of make this assessment of how you're feeling? [064-HAY, 33, F]*

Similarly, awareness of mood seemed to act as a positive reinforcement:

*I think actually I was on the happier end of it more as time went on. I think, because it did make me feel quite a bit better even if, you know, I think I went through a really stressful period at work, but actually it didn't affect me like it might have done, I would say. I think I might have ended up unwell and off work with the amount of work I ended up having to do. But I didn't. I actually flew through it, kind of thing. So yes, for me, that was quite an interesting exercise. [034-SAR, 34, F]*

## **The type of prompting**

It has to be noted, that some participants enjoyed the barking (PACO's default prompt that delivered the EMA) as the prompt to engage with the app. Some participants, however, had more negative feelings about the sound. These to extract illustrate the mixed feelings about the notification:

*I thought it was actually quite funny. I think the choice of the dog was a good one because the dog is a popular animal. Most people like them. I like dogs, I like cats as well. I'm not one of these people that.*  
[031-MAR, 53, M]

*And:*

*The barking, annoyed, violent – so I normally have my phone on, on vibrate or silent. But it vibrates my watch when the, when the dog wanted attention. [052-WAR, 44, M]*

## **Technical issues with the app software**

As mentioned in the previous chapter (section 'Ecological Momentary Assessment of mood'), there was an issue with the automatic notification on Android phones. In addition, a handful of participants would receive notifications in the night:

*Yes, so I just – the email thing was like, I don't understand why. I can understand why it might be my email, but not my email's password. And yes, just like the barking made me – like I'd be sitting at my desk and it would be going (makes a barking sound) and I'd be like, "Oh." Not even like 'woof'. It was like (loud barking noise). And yes, and it just would do it in the middle of the night. [023-ADR, 32, F]*

#### **6.4.1.4 Follow up questionnaires**

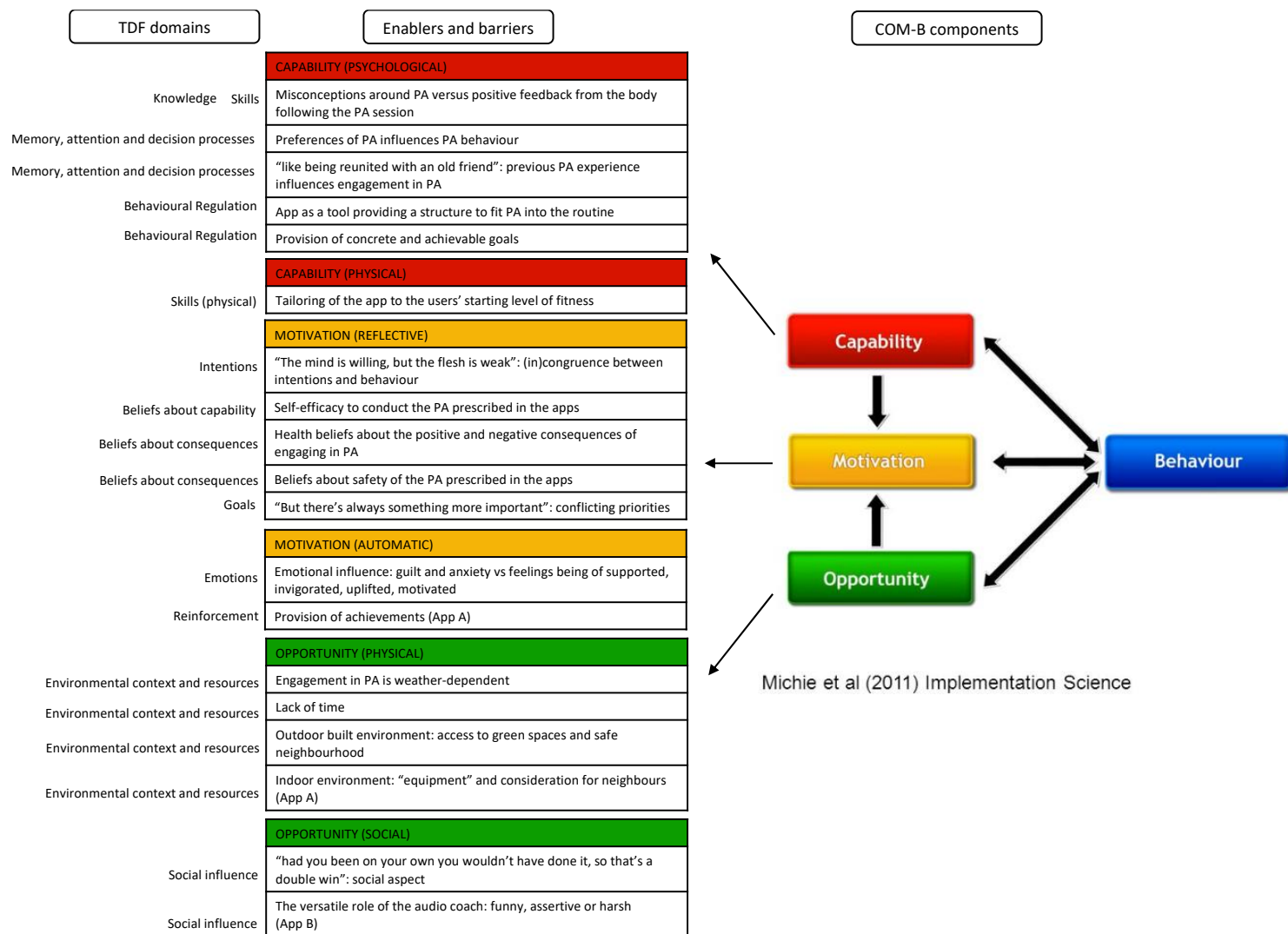
None of the participants had any issues with the questionnaires. Some participants mentioned their increased awareness when filling in the IPAQ which asked about the time (hours and/or minutes) of engagement in PA. For example, the realisation of how much time is spent sedentary:

*Yes, yes, I've done or how – I mean it was crazy for me to just realise how much I sit. There is that question, how many hours a day you sit. And really, like being super honest, fourteen – that was really crazy, you know, it was just crazy to realise that that is the reality. [028-ADA, 30, F]*

#### **6.4.2 The influences (enablers and barriers) on PA behaviour and how apps can help or hinder PA engagement (their potential as a behaviour change tool)**

The map summarising the themes within the model of behaviour (COM-B) with TDF domains is presented in Figure 55.

Figure 55: Summary of the themes with COM-B / TDF classification





The results are presented for each of the COM-B components for both behaviours tackled in this study: 1) PA in general, 2) app-specific PA behaviour, and reflect the goals of this section which were to identify the influences (enablers and barriers) of PA behaviour, and to identify how the two app interventions can help or hinder PA engagement (presented in sections: *Apps as potential behaviour change tools*).

There were instances where themes related to either PA behaviour specifically or to app PA behaviour but not both. In such cases, this is indicated in the summary table and the results. Because of the nature of the trial, some parts of the analysis are presented for each app specifically, if there were themes related to each of the apps that needed addressing.

The subsequent results are described using both COM-B and TDF, with COM-B offering a higher level summary and TDF offering a more granular level of analysis. Each section of the results begins with a table summary of the COM-B components: Motivation (Table 65), Opportunity (Table 66), and Capability (Table 67).

#### **6.4.2.1 Motivation**

This section of the results presents the influences on behaviour classed within the Motivation domain of the COM-B model. The summary of the themes representing influences on PA behaviour in general (i.e., enablers, barriers or both) and app-specific PA behaviour (as potentially enabling or hindering engagement in PA), with example quotes is presented below. Following, each theme is presented in-depth with accompanying supporting quotes.

Table 65: Summary table of the themes with TDF/COM-B classification for the domain Motivation

COM-B	TDF domain	Enabler, Barrier, Both	Theme – influences on PA behaviour in general	PA behaviour in general Example quote(s) [Study ID, age, gender]	Apps as potential behaviour change tools	App-specific PA behaviour Example quote(s) [Study ID, age, gender]
REFLECTIVE MOTIVATION	Intentions	Both	“The mind is willing, but the flesh is weak”: (in)congruence between intentions and behaviour	<i>I guess of those cases where the mind is willing, but the flesh is weak. So, it's like I know I should do it, I know it's good for me, but I just don't want to.</i> [064-HAY, 33, F]	<b>Enabler</b> through raising awareness of own PA and the perceived improvement following completion of the PA session	<i>I thought, “Oh my God, if it's going to be like this every time I do it, I'm not sure that I'll keep it up.” And doing it the second, the third and the fourth time, it just got easier and easier. And I think that makes a difference in terms of your commitment to do it. If I'd done it and it had been as challenging every time and I hadn't noticed any improvement, then I think I would have struggled to keep it up.</i> [054-NIG, 53, M]
	Beliefs about capability	Both	Self-efficacy to conduct the PA prescribed in the apps	<i>So I guess you would find out, because I do know people who have used this, and they said they loved it. But they're also built very slight and, you know, kind of natural runners, whereas I'm not.</i> [052-WAR, 44, M]	<b>Enabler</b> through the ability to go at own pace using the app and the positive feedback from the body following completion of the PA session; it may <b>hinder</b> if the user perceives their inability to do the PA targeted in the app	<i>I found it quite easy to sort of keep up with the exercises. Because I could pause it when I wanted to and there was also a demonstration as well of a man and a woman doing the exercise and so it was easy to know as to what, how to perform them. And that gave me more confidence as well in doing it accurately.</i> [025-OUS, 25, M]  <i>So, I played around with that, but I found when I was playing around with it (...). And I, like my ab muscles are like non-existent. So, I know it's something I need to work on but working on it isn't like, it's hard to do ab workouts.</i> [010-LUC, 19, F]

	Beliefs about consequences	Both	Health beliefs about the positive and negative consequences of engaging in PA	<p><i>And I think there's something about, generally speaking, I want to be, you know, I want to be healthy. I don't want to be Type 2 diabetic and I don't want to have high blood pressure or I don't want to drop dead of a heart attack. So there's something about just wanting to be comfortable, healthy really. [054-NIG, 53, M]</i></p>	<p><b>Enabler</b> through engagement with the app, and the voice coach; may <b>hinder</b> if the PA was perceived as potentially detrimental to health or not congruent with health goals</p>	<p><i>Yes, that was an early morning run. 'I'm working better with the coach.' Yes, it was good, because I woke up and I hadn't done the run the day before, or maybe it was like two days and I thought, "Oh I have to do one today." And that was about seven in the morning. I couldn't sleep. So I hadn't had a good sleep so I thought, "Right I'll just go out and try and clear my head," and it was really good. So I did that. [006-CAT, 26, F]</i></p> <p><i>I think I'm open-minded about it. But I'm not sort of desperate to sort of do that do that kind of exercise. Because I'm not sort of overweight or obese that would require me to do that. [025-OUS, 25, M]</i></p>
--	----------------------------	------	---	--	---	--

	Beliefs about consequences	Both (app-specific)	Beliefs about safety of the PA prescribed in the apps	N/A	<b>Barrier:</b> Safety issues with no thorough guidance of how to conduct the exercises safely and lack of warmup/cooling down (App A); safety of running in urban spaces and the potential for injury (App B); <b>enabler:</b> Guidance and incremental increase in running was seen as safe (App B)	<i>I think I was, I was a bit, not really concerned – I guess I was surprised the way it takes you straight into the exercises with no warmup. And that's probably not so much of an issue now that I'm finding it a bit easier. But at the start, I thought, "Oh my God." (...) And it didn't ask you to do any stretches at the end. So I was a bit surprised about that. [054-NIG, 53, M, App A]</i>
	Goals	Barrier	"But there's always something more important": conflicting priorities	<i>It's like one of the things that it's good to value exercising and I'd like to get somewhere where I'd like to value it but it's not happening. I think, yes. [010-LUC, 19, F]</i>	Not mentioned explicitly how apps can help with the prioritisation of PA; no app content features that enable prioritisation	N/A

AUTOMATIC MOTIVATION	Emotions	Both	Emotional influence: guilt and anxiety vs feelings of being supported, invigorated, uplifted, motivated	<i>I feel so guilty, honestly. Everyday. But then, I don't know, just life gets in the way I suppose.</i> [006-CAT, 26, F]	<b>Enabler</b> through evoking positive emotions through engagement with the PA and the feature of the voice coach; may <b>hinder</b> when the PA is too difficult or when the repetitive nature of the PA is seen as boring	<i>Yes, so I was surprised how good I felt. And it is, I think it invigorates you, I think there's something about a bit of clarity of mind of, you know, because I'm the sort of person that thinks a lot and it's quite good, you know, just to go out and run.</i> [054-NIG, 53, M]  <i>I think it's they're boring and I think, I don't know, like standing up on a chair, I feel, "Is that safe?" I'm not very coordinated and like a chair is like a third of my height, and it's kind of high to stand of that.</i> [023-ADR, 32, F]
	Reinforcement	Enabler (app-specific)	Provision of achievements (App A)	N/A	<b>Enabler</b> through providing sense of achievement and the motivation to collect the rewards	<i>I think it just really works as a reward, and then your dopamine system goes like someone has just told me 'well done', you know. And that is, and that adds to that overall satisfaction that I've, I have got some achievement.</i> [028-ADA, 30, F, App A]

#### 6.4.2.1.1 Reflective Motivation

##### ***“The mind is willing, but the flesh is weak”:* (in)congruence between intentions and behaviour (Intentions)**

Many participants, following the download of the app, described conducting some mental planning around the use of the apps. Many participants reported that the act of downloading the app interventions increased their awareness of their PA or lack of it. However, it was clear that, for some, simply raising awareness was insufficient to overcome barriers, as can be seen in the quotes below:

A meaningful quote summarising the phenomenon around the discrepancy between the intention and the behaviour is illustrated by participant expressing their honest recollection of why they have not engaged in using the apps to increase their PA:

*I guess of those cases where the mind is willing, but the flesh is weak. So, it's like I know I should do it, I know it's good for me, but I just don't want to. [064-HAY, 33, F]*

Similarly, 045-LIA quote reflects the triggering of thinking about engaging although not increasing PA in the context of the study:

*I think the whole idea of it was good. I mean I – it did make me think how I could fit running into my daily routine. So it got me thinking, but not necessarily doing it. [045-LIA, 26, M]*

Some participants talked about the app interventions as triggers to make them think about exercise. For example, participant below [023-ADR] seemed to increase their intentions to do PA using the app a which seem to be meaningful for them:

*I mean, you know, if you think about it, participating in this has raised my awareness of opportunities to fit in working out. I've actually worked*

*out a little bit and I even started the Couch to 5K thing. So like, just because I didn't really show a lot of progress, to me this is a lot of progress. [023-ADR, 32, F]*

### **Apps as potential behaviour change tools**

However, some participants talked about the app interventions as triggers to make them think about exercise, and engage in the behaviour. For example, participant 023-ADR seem to increase their intentions to do PA using the app which seemed to be meaningful for them:

*I mean, you know, if you think about it, participating in this has raised my awareness of opportunities to fit in working out. I've actually worked out a little bit and I even started the Couch to 5K thing. So like, just because I didn't really show a lot of progress, to me this is a lot of progress. [023-ADR, 32, F]*

Many participants talked about the process of deciding on the frequency of their PA using the apps, following downloading the app. The abstract below reflects the intentions to increase PA but, due to various factors, the reality of participants conducting PA using the apps was different from what they had planned.

Those that engaged with the PA behaviour using the apps spoke about the planning similarly to those that did not engage however, the planning was "knowing" and acting:

*So I'd go to bed knowing that I was going to do it the following morning. And that was enough for me to wake up early and I was, you know, was kind of awake by five o'clock, dragged myself out of bed by 05.30, quarter to six or whatever, and was out for half an hour. And it's fine. [054-NIG, 53, M]*

Similarly, to increasing confidence (discussed in the next theme) engagement in PA behaviour through using apps increased intentions to conduct PA and this, in turn, facilitated further engagement:

*I thought, "Oh my God, if it's going to be like this every time I do it, I'm not sure that I'll keep it up." And doing it the second, the third and the fourth time, it just got easier and easier. And I think that makes a difference in terms of your commitment to do it. If I'd done it and it had been as challenging every time and I hadn't noticed any improvement, then I think I would have struggled to keep it up. [054-NIG, 53, M]*

#### **Self-efficacy to conduct the PA prescribed in the apps (Beliefs about capability)**

Participants spoke about their confidence and its impact on the engagement in the exercise prescribed in the apps. Some of them did not believe in their stamina to conduct certain exercises, others believed in their innate lack of the ability to run "well". Some participants mentioned the apps helping with their perceived ability to engage in PA and, in turn, the sheer engagement in the exercise increased their self-efficacy. In addition, the demonstration of the exercise and the ability to go at the users' own pace acted as a facilitator.

This quote reflects the impact of low self-efficacy, which results in postponing of the behaviour: "I'll do this tomorrow":

*I'm doing it, it's like I get very (sighs) like that, and then I feel like, kind of like bad and guilty for getting like that because I've only done a little bit and it doesn't make any sense. So when I'm trying to push myself, I feel a bit weak in myself, "I'll do this tomorrow." [004-FIO, 30, F]*



Some of the participants seemed to believe in their inability to engage in certain types of PA and perceived it as a stable identity or a trait feature rather than a skill that could be practiced and improved:

*So I guess you would find out, because I do know people who have used this, and they said they loved it. But they're also built very slight and, you know, kind of natural runners, whereas I'm not. [052-WAR, 44, M]*

And further, the potential for the negative social consequences of running in public as a consequence of this perceived *inability* to run well:

*You know, I've got a funny run, so I run funny, so people are going to look at me, they're going to think, oh, you know, I'm no good at it. You know, it's all the stuff. I guess, I don't know, I'm talking through it now, but this is not the sort of thing that's necessarily at the forefront of my mind. But I do think it's there. I think that these things come from a long time ago. And they still have an impact now. And make it difficult to get going. [052-WAR, 44, M]*

### **Apps as potential behaviour change tools**

This participant talks about the content of the app as increasing their confidence in doing PA, specifically the demonstration of the exercise and the ability to pause the workout session if needed:

*I found it quite easy to sort of keep up with the exercises. Because I could pause it when I wanted to and there was also a demonstration as well of a man and a woman doing the exercise and so it was easy to know as to what, how to perform them. And that gave me more confidence as well in doing it accurately. [025-OUS, 25, M]*

In relation to the app use, another participant mentioned their engagement with the app and how the feedback from the body enabled them to develop more confidence in their ability to conduct the exercise:

*I thought, "Oh my God, if it's going to be like this every time I do it, I'm not sure that I'll keep it up." And doing it the second, the third and the fourth time, it just got easier and easier. And I think that makes a difference in terms of your commitment to do it. If I'd done it and it had been as challenging every time and I hadn't noticed any improvement, then I think I would have struggled to keep it up. [054-NIG, 53, M]*

The apps may hinder PA if the user perceives their inability to do the PA targeted in the app. An example of the participant having the ability to believe in their muscle strength or stamina to engage in certain exercise from the PA session:

*So, I played around with that, but I found when I was playing around with it, most of the workouts here they're like ab workouts. Or like leg workouts. And I, like my ab muscles are like non-existent. So, I know it's something I need to work on but working on it isn't like, it's hard to do ab workouts. [010-LUC, 19, F]*

### ***Health beliefs about the positive and negative consequences of engaging in PA (Beliefs about consequences)***

There were various beliefs that participants had about their engagement in PA that acted as both enablers and barriers to increasing PA. In particular, some participants believed that it might not be possible to influence their health in a meaningful way due to forces beyond their control, whilst others strongly believed that they could preserve their health through engagement in PA. Others mentioned more immediate results of PA, for

example, helping with clearing their mind before engaging with their work, or coping with stress.

There were also specific views around the PA targeted in the apps. Some participants believe that the PA targeted in the app may be detrimental to health, as in the case of running. Some participants believed that certain types of PA are appropriate for weight loss and, hence, not congruent with their health goals.

An example of the perceived positive effect of PA is illustrated in this participant's quote:

*And I think there's something about, generally speaking, I want to be, you know, I want to be healthy. I don't want to be Type 2 diabetic and I don't want to have high blood pressure or I don't want to drop dead of a heart attack. So there's something about just wanting to be comfortable, healthy really. [054-NIG, 53, M]*

The participant's view above is juxtaposed with the belief that unpredictability of events means that engagement in health behaviours might not be salient for them:

*Because if you think about it like that. I've had friends that are very healthy, and they've had strokes and everything. So, it's not. I don't know. It's not about, you know, it's about you, isn't it? As a person your body if how it reacts to whatever it reacts to when you're older. [056-CAR, 42, M]*

Some participants perceived that they did not experience the benefits of PA, either immediate or long term. For example, participant WAR spoke about their received like a benefit although they have been extremely motivated to engage in PA using the service of a personal trainer. This theme is closely linked to the role of emotions towards PA (theme 'Emotional influence'). For them PA seemed to be a constant struggle:

*I had personal training on a Friday and then I went to a gym on a Saturday and Sunday. And I did that for about six months at a stretch. And even though my instructor was informing me that I was getting stronger (...). But I didn't feel any better. I felt just as bad in six months as I did in the first month. And I didn't like it, just don't like it. It's just how I am. So, you know. You know, people always say, "Oh, you know, it releases endorphins, you feel so much better." I don't disbelieve them, I'm sure it's true. But I've never experienced that. The only thing I've ever experienced is pain and exhaustion basically. [052-WAR, 44, M]*

### **Apps as potential behaviour change tools**

On the contrary, this participant believed in the positive impact of PA on mental health and that the app helped them to facilitate their PA engagement via the voice coach:

*Yes, that was an early morning run. 'I'm working better with the coach.'*  
*Yes, it was good, because I woke up and I hadn't done the run the day before, or maybe it was like two days and I thought, "Oh I have to do one today." And that was about seven in the morning. I couldn't sleep. So I hadn't had a good sleep so I thought, "Right I'll just go out and try and clear my head," and it was really good. So I did that. [006-CAT, 26, F]*

Some participants believed that even a minimal engagement with the exercise, which was facilitated by the app, would be beneficial for them despite not engaging with the while workout routine:

*Yes. I know I didn't use it the whole time, but like it really got me thinking about exercising. And like I've done a few minutes of exercise*

*in the last few weeks. I think I've done like, I don't know, maybe two minutes a day in the last few days. So that's cool, better than nothing. And yes, I think maybe not for what it was trying to get me to do, but I think it was still effective in getting me to do things. [023-ADR, 32, F]*

Certain beliefs about the function of PA would stop some participants to engage in PA, as in the case of this participant:

*I think I'm open-minded about it. But I'm not sort of desperate to sort of do that do that kind of exercise. Because I'm not sort of overweight or obese that would require me to do that. [025-OUS, 25, M]*

Some participants also held beliefs that the PA targeted in one of the apps might not be healthy:

*The other thing is that I feel that I don't believe running's actually that good for you, because I believe it harms the joints of your legs (...) I'm sure there are injuries from people who run too hard and beyond their abilities and I don't want to hurt myself in that way, because I already have strained ankles and they're very painful. [031-MAR, 53, M]*

***Beliefs about safety of the PA prescribed in the apps (Beliefs about consequences)***

Concerns about the safety of the content of the apps in relation to the PA sessions were expressed by many participants. Different issues for each of the apps were raised and hence are discussed separately for each app interventions.

## **Apps as potential behaviour change tools**

### **App A**

For App A, the lack of information and guidance about conducting the various exercises was concerning, as well as the lack of warmup or stretches routine before and after the PA session.

Firstly, the safety of the information provided that would enable participants to engage in the exercises provided in apps were questioned. As such, there might be concerns that participants might injure themselves. Some participants dealt with it by looking for more information about how to do the exercise safely, for example:

*But what I actually found was that I had to do more than what was on this app to be able to do it. So, before starting, I thought I actually am ashamed to say I didn't know how to do a press-up. And there's, there's directions on here, but I felt like I needed more than that. So I had to go on to YouTube and do like a series of stages to be able to do a press-up, if you know what I mean. I didn't just go straight into doing it.*  
[034-SAR, 34, F]

Of interest, the lack of knowledge made the participant “ashamed” as they might have perceived that the norm is that the user knows how to engage in PA safely (the importance of emotions in PA is summarised in theme: Emotional influence). However, no alternations based on the fitness level were provided. It had to be noted that when the audio guide was changed to female, there was an altered, simpler version of the press-up. However, there might be an assumption about the gender differences in doing certain exercises.

Secondly, participants expressed a concern about the lack of the warm-up and stretch routines. For example:

*I think I was, I was a bit, not really concerned – I guess I was surprised the way it takes you straight into the exercises with no warmup. And that's probably not so much of an issue now that I'm finding it a bit easier. But at the start, I thought, "Oh my God." (...) And it didn't ask you to do any stretches at the end. So I was a bit surprised about that.*

[054-NIG, 53, M]

## **App B**

The safety issues that were expressed by participants for App B were concerns about running in urban spaces and the potential injury from running. At the same time, some participants appreciated the running programme, which might be perceived as a guide to injury-free PA engagement.

The main issue was the concern about running outside in urban spaces because of the traffic and running unaccompanied:

*I think outside is a little bit dangerous sometimes with traffic, with. Things happen to people. So, I think you're exposing yourself to some dangers (...) I don't know how single females feel about running early hours of the morning or after work at night in the winter time, especially when it's very dark. I don't feel comfortable or safe doing it. [031-MAR, 53, M]*

Although App B includes information about safety, it was rarely mentioned by the participants. Below is an example of participant that read the safety information but they themselves doubt that other users would:

*Yes, so it was good before you start. Yes, actually all this stuff is something I haven't looked at, but 'make sure you have comfortable running shoes, supported bra. Remember to carry your device with you*

*and so that you can listen to the trainer. Some people use an arm band.' So it is all there. I suppose what might have been helpful (...) is to say, you know, 'Read the safety information,' or 'Read everything.' Maybe it did say that and I missed it. But, you know. [034-SAR, 34, F]*

At the same time, Participant 028- ADA described their "oh this is too easy for me" attitude and, following, their knee pain which they attributed to not following the running programme. This seemed to be a learning point for them that the app facilitated:

*But as I was running (...) I had kind of like 'oh this is too easy for me' kind of attitude, it was nice to notice this. And it was nice to kind of like get the knee pain at some point just thinking, 'no, you don't know better,' you know what I mean. I don't really know better. And it's nice that there is an app and then, you know, and coaches who have to run 5K without any kind of – I believe, without any injury or something. So, so yes, that was one lesson. So I think I went humble and then the second week, how I was supposed to, which was nice. [028-ADA, 30, F]*

Of interest, one participant mentioned the data safety in the context of using the apps to connect with others:

*I don't like connecting to other people, especially other people I don't know. I spent some time in the cyber security world. Yes, I don't like connecting much to anything. I really don't. There are so many vulnerabilities, so many things out there and it's just getting really bad and it's only going to get worse. [066-CHU, 48, M]*



***“But there’s always something more important”: conflicting priorities (Goals)***

The role and value placed on deliberate engagement in exercise is difficult with many conflicting life priorities. These included family, friends, work and career progression. Hence, the daily activities revolved around these aspects of life. As the population of interest in this study were people residing/working in urban space, the nature of work would be predominantly sedentary, the emphasis placed on the importance of work, and career development meant that PA engagement featured low in the ranking.

The quote below reflects this well the positioning of PA on the list of priorities:

*It’s like one of the things that it’s good to value exercising and I’d like to get somewhere where I’d like to value it but it’s not happening. I think, yes. [010-LUC, 19, F]*

025-OUS talks about prioritising the family and the importance of skills development to progress in the career:

*Helping my friends and family. Yes, so yes, they have their own sort of things that they want to do. So, socialising as well. And basically, learning new skills and technologies and so on. Yes, because they help me in my career. And so, yes, those would be some of the reasons, I think. [025-OUS, 25, M]*

Prioritising the needs of others first featured, for example, the experience of being a parent and ensuring the activity needs of the children are met first. This leaves little time for caring for the self:

*My life. My kids. My kids are my priority. Because my kids are very active. They’re really active kids. They do a lot of football, I’m very football orientated. I’m a season ticket holder as well at Chelsea. So, it’s kind of my life focuses a lot on football. I don’t do it, like constant*

*like. I would love to do it as a career, but I don't. But I love that type of thing. So, it's, that's my kind of main focus. [056-CAR, 42, M]*

As other life domains would fight for attention, they would create a sense of conflict and, as in the case of the participant's response below, a sense of guilt:

*It's a dilemma because the thing I'm doing now is also something that's very important to me. So going to my course and work is definitely taking me in the direction I want my life to go as is going to the gym. If I was to go to the gym and it's happened a few times where I've gone to the gym and thought I need to be at home writing. I've got a case report to write up. It's getting in the way. [022-MIC, 36, M]*

Similarly, participant 064-HAY expressed their conflict around their engagement with the behaviour and the effortful, deliberate consideration of what priorities should be pursued:

*But if it's a reminder that I need to set for myself again it's like making a plan to say, "Okay, I'm going to run." (...) But there's always something more important. Yes. So, it's quite easy to shut that down. And be like yes, "No, I have to, you know finish this essay." Or you know, it's like, "No, I'm really tired. I just want to take a shower and go to bed." [064-HAY, 33, F]*

### **Apps as potential behaviour change tools**

None of the app content features enabled prioritisation to make PA more salient in day-to-day life. However, the next theme provides an insight into how apps can provide a sense of structure that might facilitate the prioritisation of PA.

#### 6.4.2.1.2 Automatic Motivation

##### ***Emotional influence: guilt and anxiety vs feelings of being supported, invigorated, uplifted, motivated (Emotions)***

Affective states around engaging in PA and the use of apps were prevalent in participants' responses. These acted as either enablers or barriers to engagement in PA. First, the results of negative emotions around non-engaging in PA itself is summarised, following by both positive and negative emotions connected to the use of the apps in this study.

Participants described the negative impact of their non-engagement in PA as experiencing guilt and low mood. However, adhering to an exercise routine that was not enjoyed also lead to negative feelings.

The quote below illustrates the sense of guilt when participants did not meet their own expectations and consequently procrastinated regarding their engagement in PA:

*I feel so guilty, honestly. Everyday. But then, I don't know, just life gets in the way I suppose. [006-CAT, 26, F]*

Conflict around commitments, such as family or career (described in theme "But there's always something more important": conflicting priorities) often results and low mood was another barrier for the participants creating a vicious circle of *low-mood-PA-non-engagement-low-mood*:

*I come back, I'm sad. I feel that I wasn't really productive as I want me to be. And then I know that if you go for running or if you do exercise, it's going to help you. But you feel so low that you won't go that step (...) I know that if you do exercise a good way to release your stress, release your thoughts but it doesn't always work. [042-MAR, 25, F]*

Shame was a strong reaction of participant 052-WAR describing the process of seeing personal trainer with whom they engage for six months at a time in order to increase their fitness:

*I also find that – and I know this is a bit off topic, but I find it humiliating, a lot of the stuff that the personal trainers get you to do. I just don't like it (...). I feel like, a) I'm not very good at it. And then b) I'm being shouted at by somebody to do it, to go faster and harder and all that kind of thing. And I'm feeling extremely self-conscious. [052-WAR, 44, M]*

### **Apps as potential behaviour change tools**

When participants engaged in PA that was enjoyed, they described feeling supported, invigorated, uplifted, and motivated. However, some of the participants who liked the app, did not necessarily increase their PA. There were also some negative emotions experienced, such as boredom around the lack of variety in the prescribed PA routines offered by the app. For those participants that found some of the exercises difficult, it also leads to negative emotions.

### **App-related positive emotions acting as enablers to PA**

The voice coach in App B seems to evoke emotions of feeling positive and supportive, and motivating for using the app:

*She's not very, you know, that, for example, I had a personal trainer in the past. He was very like military style, and I don't like that. And when I listen to her, she doesn't have that tone in her voice. She's more like, I don't know, a supportive peer. [006-CAT, 26, F, App B]*

And:

*'I'm working better with the coach.' [006-CAT, 26, F, App B]*

Participant 054-NIG described the “uplifting” and “motivating” feelings after completing the running session with the app, both on the body but also on the “clarity of mind”. In turn, the positive feeling increased the engagement with the consecutive running session:

*Yes, so I was surprised how good I felt. And it is, I think it invigorates you, I think there's something about a bit of clarity of mind of, you know, because I'm the sort of person that thinks a lot and it's quite good, you know, just to go out and run. [054-NIG, 53, M]*

However, for other participants, the positive emotions about the apps were not necessarily followed by engagement with the exercise routine as in the example of participant 023-ADR:

*Yes, I really liked it even though I didn't ever do seven minutes. [023-ADR, 32, F]*

Engagement with the app, together with the physiological reaction of the body, created a sense of the “feel-good”:

*For me, the actual – so the exercise was, so yes, these are the exercises. I did it, the first day I had it and I was sweating like there was no tomorrow, afterwards. I felt so good doing it, I was so tired. [045-LIA, 26, M]*

### **App-related negative emotions acting as hinderers to PA**

The repetitive nature of the apps was a barrier to using of the apps as the participants often talked about boredom in connection to these routines. The two extracts below summarise the repetitive nature of the routines.

*I think it's they're boring and I think, I don't know, like standing up on a chair. [023-ADR, 32, F]*

Similarly,

*(...) by the end of that week, that second week, when I was doing two cycles each time, then was the time when I kind of like ran, slightly kind of got bored with it – I could do a little bit of a variety with, with everything. [028-ADA, 30, F]*

As well as boredom, other negative emotions could be evoked regarding frustration at not being able to complete the exercises. The following extract illustrates the importance of setting achievable and tailored goals in order to increase engagement as the inability to perform some of the exercises evoked the feelings of frustration and annoyance:

This extract illustrates the importance of the achievable goals that were tailored to the skills of participants in order to increase engagement as the inability to perform some of the exercises evoked the feelings of frustration and annoyance:

*So it was a bit annoying, because like I could do some of them quite easily, like the jumping and everything. And then you got to one and you would sort of like try but you felt like you weren't managing it. So it probably just wasn't quite the right type of things (...) like I like to go at my own pace when I do exercising and when I do things. And then like it felt too hard quite a lot of the time, and that was a bit frustrating. And it, and then, even though it's like really short and meant to be quite intense and then you get to rest and it sounds like it's quite low impact – it felt quite hard, I don't know why, I found it quite – it's probably quite hard to do. (...) it was a bit demoralising. [063-MAR, 26, F]*

### ***Provision of achievements (App A, Reinforcement)***

The gamification element was a feature in App A hence this section relates to the 7 minute HIIT exercise. The provision of rewards seem to evoke a sense of satisfaction that followed the achievements of certain rewards. The rewards also enabled the participants to self-reflect on what they had completed and provided a source of motivation. Yet, for some participants the rewards did not have any impact and were seen irrelevant to their engagement with the PA using the apps.

The satisfaction provided by the achievement of rewards was illustrated by participant 028:

*I think it just really works as a reward, and then your dopamine system goes like someone has just told me 'well done', you know. And that is, and that adds to that overall satisfaction that I've, I have got some achievement. [028-ADA, 30, App A]*

Similarly, the response of participant 042-MAR suggests that the rewards seemed to act as a motivation, with small goals to work towards, for example, an achievement related to the continued use or completing two sessions in one day. This particular participant expressed their annoyance when they did not achieve a reward due to the app treating any session after midnight as counting towards the next day's session. This might discourage some potential users:

*I was running... Sorry, I was starting the app quite late like after midnight before (...) and it showed another day. Yes, so I missed... Due to this, I missed some kinds of rewards that you keep for six days. So I missed that and it's not my fault, I did it. [042-MAR, 25, F]*

In addition, the material reward of “unlocking” an extra PA routine when a certain number of PA sessions were completed was mentioned by some participants as motivating, for example:

*I suppose I was tempted to buy them like without waiting. But then I thought there's no point in doing that because, in the end, this is what I always do with apps and that as well, I just buy all the bonus stuff and then don't use it. So I thought if I'm going to go, if I'm going to go through and get those bonuses, then I must have earned it, rather than just pay for it. [052-WAR, 44, M, App A]*

And:

*The fact that you can, you earn, you can earn it, so you could sort of pay for it. I thought it's a very ingenious like idea. [045-LIA, 26, M]*

However, the rewards did not have a motivational impact for some participants, e.g.:

*Interviewer: And were these in any way motivating? The rewards?*

*Participant: No, not particularly. [054-NIG, 53, M]*

And:

*I liked, I like when you get a little achievement. And I got a few. I don't know. But it didn't seem to... It didn't seem to bother me that much, like it wasn't... It wasn't... Yes, it wasn't particularly like exciting or good or bad. [063-MAR, 26, F]*

#### **6.4.2.2 Opportunity**

This section of the results present the influences on behaviour classed within the Opportunity domain of the COM-B model. The summary of the themes representing



influences on PA behaviour in general (i.e., enablers, barriers or both) and app-specific PA behaviour (as potentially enabling or hindering engagement in PA), with example quotes is presented below. Each theme is then presented in-depth with accompanying supporting quotes.

Table 66: Summary table of the themes with TDF/COM-B classification for the domain Opportunity

COM-B	TDF domain	Enabler, Barrier, Both	Theme – influences on PA behaviour in general	PA behaviour in general Example quote(s) [Study ID, age, gender]	Apps as potential behaviour change tools	App-specific PA behaviour Example quote(s) [Study ID, age, gender]
PHYSICAL OPPORTUNITY	Environmental context and resources	Barrier	Engagement in PA is weather-dependent	<i>Then to me weather counts a lot. So, if it's cold it demotivates me so much. Even though I can do these exercises at home I know that. But to me it's a totally factor. If it's really, really cold, as it was a few weeks ago. [006-CAT, 26, F]</i>	<b>May hinder</b> when the app tackles PA conducted outside (App B)	<i>By the time we got through and that was sort of the weekend between the week 4 and 5, by the time I got back to there, I just didn't, and I think it was pouring with rain and that's another issue. I wasn't going to go out running in the rain. [033-STE, 64, M]</i>

	Environmental context and resources	Barrier	Lack of time	<p><i>I was doing a lot of late, like, nights at work. Which in turn, in the morning, I would have to turn up to work at, like, early in the morning as well. (...) So, yes. That's why work has taken, like, a priority at the moment. But, like, probably, I would say, like, last year, work came second. Family and friends came first (...) [043-JEN, 26, F]</i></p>	<p><b>Enabler</b> as the PA prescribed in the apps was seen as fitting into the busy routine</p>	<p><i>Time is the biggest enemy for me. So my runs are typically early morning, 05.30 to 06.00. But I quite like this as I run like Forrest Gump and it stops the neighbours from laughing at me. [054-NIG, 53, M]</i></p>
	Environmental context and resources	Both	Outdoor built environment: access to green spaces and safe neighbourhood	<p><i>And she says, "Oh, you know," she did this Couch to 5K, "When I run, I go across the millennium bridge and it's really beautiful." And I always want to say, "Oh shut up," you know. It's like, I would probably run as well, if I could run over the millennium bridge. But, you know, I've got the Harlesden Railway Bridge to run across. That's the one I've got. [052-WAR, 44, M]</i></p>	<p><b>Enabler</b> thorough provision of information on how to run safely (App B)</p>	<p><i>And I jogged in – you can jog around pavements and things, it's fine, because I'm not very fast. I'm not a threat to other people. So yes, it has good advice on the thing about how you can, like what to do, where to run and stuff. That is a support thing. [063-MAR, 26, F, App B]</i></p>

	Environmental context and resources	Both (app-specific)	Indoor environment: “equipment” and consideration for neighbours (App A)	N/A	<b>Enabler</b> when perceived as convenient to conduct the PA indoors; may <b>hinder</b> when the indoor space inconvenient (e.g., lack of stable chair) (App A)	<p><i>I’m in my own privacy, I can do it, I’ve got no stresses, nobody watching what I’m doing, done it (...)</i> [045-LIA, 26, M]</p> <p><i>I don’t think I have like, I don’t have some of the, I don’t know – like, I don’t think I have a chair. I feel like - I do have a chair, but I felt it was a really high step to step on a chair, so I never did that. It was way too high.</i> [023-ADR, 32, F]</p>
SOCIAL OPPORTUNITY	Social influences	Both	“Had you been on your own you wouldn’t have done it, so that’s a double win”: social aspect	<i>So, very often in the past I sort of, not felt like going but think I’ve got to, I’m meeting [name of a friend], can’t let him down. And you get there, and you do this session and you come back and you’re feeling really good. You’re feeling it’s a double win.</i> [033-STE, 64, M]	As the apps did not provide social features, the apps may <b>hinder</b> PA engagement when PA social aspects of PA seen as important	N/A

	Social influence	Both (app-specific)	The versatile role of the audio coach: funny, assertive, harsh (App B)	N/A	<p><b>Enabler:</b> engagement with PA session if the audio coach was liked and they fulfilled the role expected by the participants; hindered if such expectations were not met; may <b>hinder</b> if perceived as un-motivating</p>	<p><i>I didn't really fancy listening to Jo Whiley. I thought Sarah Millican would be funny rather than serious. And I wanted some serious motivation, some sensible motivation. And she looked like she was a sporty type person and was probably the default one. So I just went for that. [054-NIG, 53, M]</i></p> <p><i>She wasn't really patronising (...) there's amateur and then there's like really like sort of like, I don't know, as like a proper coach to say, "Okay." And it's like, "Oh okay, yes, I want to go back to this and do it," sort of thing. You feel like there's a bit of leeway sort of thing. [004-FIO, 30, F]</i></p>
--	------------------	---------------------	--	-----	--	---

#### 6.4.2.2.1 Physical opportunity

##### ***Engagement in PA is weather-dependent (Environmental context and resources)***

The trial was conducted in the months from January until June and this period had an impact on some participants.

For some, the cold weather, independent of whether the app-related PA was to be conducted in- or outdoor, was perceived as a de-motivator to engage in PA:

*Then to me weather counts a lot. So, if it's cold it demotivates me so much. Even though I can do these exercises at home I know that. But to me it's a totally factor. If it's really, really cold, as it was a few weeks ago. [006-CAT, 26, F]*

Lastly, for some participants the motivation was present despite the unfavourable weather conditions:

*So a couple of the days I did it, it was absolute – it was horrible, the weather was horrible, it was heaving down with rain. And I don't really mind getting wet when I go out running. So, so there was no excuse about the weather. [054-NIG, 53, M]*

##### **Apps as potential behaviour change tools**

For many, the weather in winter in London posed an obstacle, especially when conducting the running using App B:

*By the time we got through and that was sort of the weekend between the week 4 and 5, by the time I got back to there, I just didn't, and I think it was pouring with rain and that's another issue. I wasn't going to go out running in the rain. [033-STE, 64, M]*

### ***Lack of time (Environmental context and resources)***

Participants often spoke about a lack of time as an obstacle to engaging in PA in the context of the trial. In the analysis, the issue with time was explored further to investigate the reasons for the perceived lack of time as this was especially important in the context of urban living. The barriers found were related to the current life priorities (discussed in the theme “But there’s always something more important”: conflicting priorities) and included long working hours, time taken to commute to work as well as family commitments. Yet, those who engaged with the apps saw them as an opportunity to fit exercise into their routines. This related to the convenience of the apps as discussed in the theme ‘App as a tool providing a structure to fit PA into the routine’.

Some participants discussed long working hours as a barrier, illustrating the reality of living in the city with competing priorities:

*I was doing a lot of late, like, nights at work. Which in turn, in the morning, I would have to turn up to work at, like, early in the morning as well. (...) So, yes. That’s why work has taken, like, a priority at the moment. But, like, probably, I would say, like, last year, work came second. Family and friends came first (...) [043-JEN, 26, F]*

Long commuting time that most likely involved passive transport was another reason for the lack of time to engage in PA:

*I think that’s not so much about time but energy. Well maybe time as well. So, I have 2 jobs, but one is part-time and the other one is really flexible as well. But they take a lot, the full day. So, one actually is in Watford. So, it takes a long time to get there. So, the travelling tires me out and the commute, it tires me out as well. So, by the time I get back it’s just, I’m just done. Yes. I don’t really want to do anything else. [064-HAY, 33, F]*

## **Apps as potential behaviour change tools**

Yet, the apps also acted as facilitators to engaging in PA in a time-constrained lifestyle. Many of the participants managed to fit some of the PA sessions as prescribed in the app into their life. They saw the app as beneficial because of this fit:

*So the apps kind of offer something for people who are really time-poor, I suppose. It makes it kind of possible to fit it in. [034-SAR, 34, F]*

Some participants whose motivation seemed to be stable despite time being a limited resource, managed to make a commitment to fit the PA session in despite their busy schedule. The following entry was from a participant's diary

*"Time is the biggest enemy for me. So my runs are typically early morning, 05.30 to 06.00. But I quite like this as I run like Forrest Gump and it stops the neighbours from laughing at me." [054-NIG, 53, M]*

## ***Outdoor built environment: access to green spaces and safe neighbourhood (Environmental context and resources)***

In this section, the importance of the environment (whether built outdoor environment or living arrangement) were described as the potential enablers and barriers to engaging with PA and with the PA using the apps.

Access and proximity to facilities (gym, green spaces) where one could engage in PA safely was important for many of the participants. The quote below reflects well the social inequalities in access to enjoyable built environments in London. 052-WAR describes two different environments in the city, one conducive to doing PA outside and the second, aesthetically unpleasing or potentially unsafe:

*And she says, "Oh, you know," she did this Couch to 5K, "When I run, I go across the millennium bridge and it's really beautiful." And I always*



want to say, “Oh shut up,” you know. It’s like, I would probably run as well, if I could run over the millennium bridge. But, you know, I’ve got the Harlesden Railway Bridge to run across. That’s the one I’ve got.

(....)

So there’s a lot of like – there’s a big FedEx thing there, there’s a big trucking company, there’s a food manufacturer, biscuit manufacturer and so on. It’s not the greatest area. It’s not the sort of place that you’d want to hang around outside, like you’d just go the station and then you go home. [052-WAR, 44, M]

### **Apps as potential behaviour change tools**

The participant below describes how the use of the App B increased their confidence by providing instructions around running on the pavement. This acted as an enabler to their engagement with the PA sessions through the app:

*And I jogged in – you can jog around pavements and things, it’s fine, because I’m not very fast. I’m not a threat to other people. So yes, it has good advice on the thing about how you can, like what to do, where to run and stuff. That is a support thing. [063-MAR, 26, F, App B]*

### ***Indoor environment: “equipment” and consideration for neighbours (App A, Environmental context and resources)***

### **Apps as potential behaviour change tools (App A)**

The advantage of App A, expressed by many participants, was its convenience (as described in theme ‘App as a tool providing a structure to fit PA into the routine’):

*It was, I mean genuinely, I could have done it in my room and I was like, do you know what, I'm in my own privacy, I can do it, I've got no stresses, nobody watching what I'm doing, done it (...)[045-LIA, 26, M]*

However, some participants described issues with conducting the exercise in small living spaces in areas of high population density, which illustrated well the obstacles of residing in London in the context of overpopulation and high costs of living. Some participants struggled with finding the space to do some of the exercises prescribed in App A and the quote reflects their determination:

*I have a studio flat and it's not big enough to find sufficient space to do the exercises. What space there is and also, I'm aware that there might be someone, depending on what time I'm doing it, there's somebody underneath me and if I'm bouncing up and down on my floor, it's their roof. So, I tried doing it on the landing, and that wasn't big enough and then eventually went down to the hallway on the ground floor in the middle of the day when no one was around and did the exercises there. And used the first couple of stairs for the step ups because I don't have a chair that I was confident was strong enough for me to do step-ups on. [033-STE, 64, M]*

An access to a safe chair may be assumed but it proved to be an issue for some:

*I don't think I have like, I don't have some of the, I don't know – like, I don't think I have a chair. I feel like - I do have a chair, but I felt it was a really high step to step on a chair, so I never did that. It was way too high. [023-ADR, 32, F]*

#### 6.4.2.2.2 Social opportunity

***“Had you been on your own you wouldn’t have done it, so that’s a double win”:  
social aspect (Social influences)***

Engaging in PA as a social process was a prevalent subject in participants’ responses. Sharing similar interest as engaging in PA evoked a feeling of commitment and “can’t let them down” attitude which acted as an enabler to PA. However, a close family member or partner could also act as an obstacle if they were less active hence emphasizing the importance of wider social circles. The specific influence of the social environment is discussed next.

The experience of taking part in an organised sports and exercise uniting people with similar interests meant participants felt like “being part of something” and “feeling included”. For example, this participant described their experience of going to a “club” to practice Kung-Fu where they felt accepted and at ease performing the exercise routine. It is interesting that they did not consider their weekly trainings at the club as an exercise:

*(...) And maybe it’s the fact that like it’s a club. So, there’s other people doing it with you. And I have friends who also go and it’s like you’re doing ab workouts but I’m failing, my friend is next to me also failing and the other person across the room is like doing it like so well. It’s like it’s a different atmosphere, I guess. [010-LUC, 19, F]*

Here, the same participant juxtaposes the experiences of being in a university club with going to the gym where they felt more self-conscious and judged:

*Yes, because the gym it’s like you don’t know anyone there. And I guess I’d be more likely. I like, I go to the gym when there’s like less people there. And the gym I went to is like a really small gym. And I could go to the one in the very corner and hope no one watches me.*

*But like I also hope no one watches me at the club when we're doing like exercises and stuff. But I don't know, I just the fact everyone's also doing the same thing and it's like I know I'm not very good at it. But it like doesn't matter that I'm not very good at it. And that isn't any different from the gym, but I think I know it's just different. [010-LUC, 19, F]*

This participant described their experience with PA as lonely, again stressing the importance of social aspect of PA:

*And it's one of the loneliest things in the world. Long distance running is very lonely. I used to run a long time ago, but I hate running now. [066-CHU, 48, M]*

Not wanting to let someone else down seemed to be important in increasing motivation to engage in PA. There was a certain commitment to fulfil the duty as the exercise partner or a buddy which increases motivation:

*So, very often in the past I sort of, not felt like going but think I've got to, I'm meeting [name of a friend], can't let him down. And you get there, and you do this session and you come back and you're feeling really good. You're feeling it's a double win. Not only you've done the exercise so that's a win, but the thing is you knew that had you been on your own you wouldn't have done it, so that's a double win. [033-STE, 64, M]*

The same participant mentions further the impact of their partner as a hindrance to them being active stressing the importance of developing other social avenues that facilitate PA engagement:

*And I thought that I'd try to use this to try and kick start myself into becoming more active again. But I just haven't managed it. And I think partly it's because of lifestyle and the influence of what I do with my girlfriend. She's got her own minor mobility issues with a sore hip that means she can't walk long. [033-STE, 64, M]*

### **Apps as potential behaviour change tools**

In the context of the app use in the trial, the extract from the participant below talks about conducting the exercise routine using App A which is a lonely and “sordid” and a juxtaposition to the social experience of doing exercise:

*There's something a bit sordid that you would lock yourself in the room to do the... I think it's the same with mindfulness stuff as well. There's lots of apps out there with meditation and stuff but it's that idea that you go in a room and close the door and do your dirty business. Jumping up and down doing star jumps looking at a video of a guy jumping up and down. Yes, it's definitely a bit weird.*

*(...)*

*These are things that you wouldn't get by just turning up to a gym and putting your headphones in and running on a treadmill. [022-MIC, 36, M]*

### ***The versatile role of the audio coach: funny, assertive, harsh (App B, Social influences)***

#### **Apps as potential behaviour change tools (App B)**

There was great variability in participants' responses the audio coach in relation to the importance they attributed to their choice of the coach, the frequency of the speech, and the role the audio coach should take on. This variability revealed that the feature of the

audio coach could both help and hinder participants' engagement in PA. The adjectives used to describe the guide varied, e.g., "funny", "supportive", "upbeat", "assertive", "harsh", "strong", "encouraging":

*All of them. I realised in the first minute that they all say the same thing, it's just different colour of the voice. And then I was okay, I'm going to go with Michael Johnson because he's also an athlete and quite strong and encouraging so I just choose this guy and I didn't really change after that. Because his colour of the voice, it's very serious. It's fun.*  
[042-MAR, 25, F]

As seen in the previous extract and the quote below, participants had expectations based on the prior knowledge of the public role of the person acting as the audio coach. One participant pointed that there was no option to "test" the voice of the audio coach, which seemed important. If no prior knowledge of the audio coach existed, participants formed their beliefs based on the images provided in App B, for example ADR-023 spoke about their thinking process behind selecting the audio coach:

*Yes, I just wanted to hear their voice and not like – I don't want to start a run until I'm going to start a run. So I wanted to hear their voice first.*

(...)

*I think it just says, "Choose." Is she really going to make me laugh? I don't know. I'm like – so he's going to make me laugh too, I don't believe it. And who are these people? And if they are really going to make me laugh, like maybe I could – like this guy was so serious. Oh, he is serious, okay. Maybe I could have like a funny person. Anyway, yes, so I kind of got hung up on that. [023-ADR, 32, F]*

Yet, some participants preferred someone “serious” to increase their motivation, for example:

*So I knew Jo Whiley and I knew Sarah Millican. And I knew Michael Johnson. I didn't really fancy listening to Jo Whiley. I thought Sarah Millican wouldn't be – I thought it would be funny rather than serious. And I wanted some serious motivation, some sensible motivation. And she looked like she was a sporty type person and was probably the default one. So I just went for that. [054-NIG, 53, M]*

For others, supportive voice was crucial:

*I had a personal trainer in the past. He was very like military style, and I don't like that. And when I listen to her, she doesn't have that tone in her voice. She's more like, I don't know, a supportive peer. I would say so. (...) I prefer someone that just motivates me, and, you know, kind of is on the same level as I am. So, I think that's why I chose her and that's why it motivates me, keeps me using the app. [006-CAT, 26, F]*

However, for others, the audio coach was too lenient or even “patronising”. 004-FIO explains that her audio voice lacked authority and was not motivating to engage in the running session:

*She wasn't really patronising but it was just like it was coming from a very sort of like – there's amateur and then there's like really like sort of like, I don't know, as like a proper coach to say, “Okay.” And it's like, “Oh okay, yes, I want to go back to this and do it,” sort of thing. You feel like there's a bit of leeway sort of thing. [004-FIO, 30, F]*

### **6.4.2.3 Capability**

The section of the results present the influences on behaviour classed within the Capability domain of the COM-B model. The summary of the themes representing influences on PA behaviour in general (i.e., enablers, barriers or both) and app-specific PA behaviour (as potentially enabling or hindering engagement in PA), with example quotes is presented below. Each theme is then presented in-depth with accompanying supporting quotes.

Table 67: Summary table of the themes with TDF/COM-B classification for the domain Capability



COM-B	TDF domain	Enabler, Barrier, Both	Theme – influences on PA behaviour in general	PA behaviour in general Example quote(s) [Study ID, age, gender]	Apps as potential behaviour change tools	App-specific PA behaviour Example quote(s) [Study ID, age, gender]
PSYCHOLOGICAL CAPABILITY	Knowledge Skills	Both	Misconceptions around PA versus positive feedback from the body following the PA session	<i>The other thing is that I feel that I don't believe running's actually that good for you, because I believe it harms the joints of your legs. Impact, the constant impact on the knees, the ankles, the feet, the bone, you know, the cartilage and the bones and compacting against each other I believe can damage them if you run too much. [031-MAR, 53, M]</i>	<b>Enabler</b> through information in the apps and the positive feedback from the body following completion of the PA session; may <b>hinder</b> when the PA targeted by the app may be seen as unsafe or not appropriate for the participant	<i>So yes, it has good advice on the thing about how you can, like what to do, where to run and stuff. That is a support thing. (...) it says why run, so then you've got all the, like why it's good for you. And the bit which I read first was about like where to run and what to wear, because I didn't want to – I wanted to do it right. (...) Yes, this bit – 'watch out for other pedestrians' and stuff. So it's quite good. It tells you, tells you what you can do. [063-MAR, 26, F]</i>

	Memory, attention and decision processes	Both	Preferences of PA influences PA behaviour	<i>But then I'm not very into this type of thing. I'm not very, doing these things is not my - I'd rather just warm up a little bit, play football. But I'm very football orientated. It's totally different, it is. I'm not into this type of things. [056-CAR, 42, M]</i>	<b>Enabler</b> if preference is congruent with the app and may <b>hinder</b> if PA offered by the apps is incongruent with set preferences	<p><i>Yes, I love being outdoors, so then that ticks that box. You get some fresh air and you get like a physical distance from the other things you're trying to do. Whereas doing that sort of intensive exercises in the same room I'm also doing the studying in, it just gets a bit stressful. [063-MAR, 26, F, App B]</i></p> <p><i>Maybe it's just not for me, but like I mean people who like doing that kind of intensive muscle-building stuff might enjoy it more. And probably, yes, a way of mixing up the order if it isn't there already. [063-MAR, 26, F, App A]</i></p>
	Memory, attention and decision processes	Both	“Like being reunited with an old friend”: previous PA experience influences engagement in PA	<i>It's quite hard to get away from this. But those team-picking things were terrible, because I would always be the last or second last to be picked. And people would argue about which team I wouldn't be on rather than the other way round, you know, this kind of thing. So it does nothing for your self-confidence as a kid. [052-WAR, 44, M]</i>	<b>Enabler</b> when the PA offered by the apps was related to positive previous experience, may <b>hinder</b> if negative experience is evoked.	<p><i>(...) it is really good to be getting back into running again. Had forgotten how much I enjoy it and will be keeping it up after the end of the trial. Had a spot of shin splint at the end of week one. So had to get out and get a new pair of stability trainers. Got the same model as previously and they were like being reunited with an old friend, made a massive difference.” [054-NIG, 53, M]</i></p> <p><i>Yes, my experience with running is always been just like running in school, doing PE lessons and I wouldn't say it was a bad experience but it's not something that I find particularly pleasurable. [064-HAY, 33, F]</i></p>

	Behavioural Regulation	Enabler (app-specific)	App as a tool providing a structure to fit PA into the routine	N/A	<b>Enabler</b> through the structured sessions and the prescribed programme	<i>Yes, especially with the running app because I was like, "Okay, I have one day to rest and then I am going for another run." I don't know, that kept me going. [006-CAT, 26, F, App B]</i>
	Behavioural Regulation	Enabler (app-specific)	Provision of concrete and achievable goals	N/A	<b>Enabler</b> through the time-specific prescribed PA session (7 min for App A and 30 min for App B)	<i>You have really no excuse not to do it other than yes, I'm lazy. But other than that, you really have no excuse. I was using it every day. [042-MAR, 25, F]</i>
PHYSICAL CAPABILITY	Skills (physical)	Both (app-specific)	Tailoring of the app to the users' starting level of fitness	N/A	May <b>hinder</b> if the PA prescribed by the app too challenging or too easy	<i>So, kind of like you can't screen people. 'Do you like being outdoors, do you...?' I don't know. Making it kind of – the main thing for me was that the first thing that you do is achievable and not really too difficult, because that, like – maybe I'm the only person in the world that can't do a press-up. [063-MAR, 26, F]</i>

#### 6.4.2.3.1 Psychological Capability

##### ***Misconceptions around PA versus positive feedback from the body following the PA session (Knowledge, Skills)***

Knowledge and skills around exercise was an important factor that played a role in participants' engagement in PA. As described in the theme 'Health beliefs about the positive and negative consequences of engaging in PA', participants' beliefs about PA represented a barrier to engaging in PA, such as the belief that running is an exercise predominantly for weight loss. This represent gaps in knowledge around the benefits of PA. Further issues around knowledge and skills are discussed below.

Some participants increased their knowledge though the use of apps, either by the information in the apps or though body feedback following the completed PA session. At the same time, misconceptions about PA and its impact meant that some participants were reluctant to engage in it. The quote below reflects one such that may act as a barrier to engaging in PA:

*The other thing is that I feel that I don't believe running's actually that good for you, because I believe it harms the joints of your legs. Impact, the constant impact on the knees, the ankles, the feet, the bone, you know, the cartilage and the bones and compacting against each other I believe can damage them if you run too much. [031-MAR, 53, M]*

This participant's extract is an example of attentional bias where heightened attention is paid to negative information. This bias is well-recognised and may affect engagement in PA [365]. Similarly, this extract shows how a lack of knowledge and skills, as well as confidence, means that participants may not engage in PA because of their fear of injury:

*I've had friends who have snapped their Achilles' tendon playing football. People have always tried to encourage me. You know, 'Come and play touch rugby or come and play five-a-side football.' I mean I never do it, because I'm, as I say, I've got two left feet, no coordination. So I just make like a fool on the field. I've got no, virtually no endurance. So running around, it's just I couldn't do it for more than five minutes. But I'm especially worried that I'll throw myself into it and snap something or break something. [052-WAR, 44, M]*

Lack of knowledge around safe engagement in PA, specifically the importance of warming up in this case, was an obstacle to conducting PA using App B:

*It seems like you're, it's too, it's too much warming up and not exercise. It should go straight to the point. Than warm up for 5, because I know you should warm up, but people, when they're doing this type of things they're not going to. It's a bit like the internet. If you put too much at the beginning and it comes a point it becomes boring. You just scroll, you don't scroll down, you just click to the next one. And that's the point – I think it should be more, right, warm up for three minutes, do the run and then just relax and drink water or something like that, not cool down for 5 minutes. [056-CAR, 42, M, App B]*

### **Apps as potential behaviour change tools**

With the use of the app, this participant increased their awareness of the importance of stable incremental increases in PA:

*But as I was running, I think I mentioned, this already like – that even that, the fact that, the fact that I had kind of like 'oh this is too easy for*

*me' kind of attitude, it was nice to notice this. And it was nice to kind of like get the knee pain at some point just thinking, 'no, you don't know better,' you know what I mean. I don't really know better. And it's nice that there is an app and then, you know, and coaches who have to run 5K without any kind of – I believe, without any injury or something. So, so yes, that was one lesson. So I think I went humble and then the second week, how I was supposed to, which was nice. [028-ADA, 30, F]*

The extract from the participant below explains how, through the information in App B, they gained some valuable tips around running through the streets which, in turn, provided some confidence and increased their engagement in running using the app.

*So yes, it has good advice on the thing about how you can, like what to do, where to run and stuff. That is a support thing. (...) it says why run, so then you've got all the, like why it's good for you. And the bit which I read first was about like where to run and what to wear, because I didn't want to – I wanted to do it right. (...)Yes, this bit – 'watch out for other pedestrians' and stuff. So it's quite good. It tells you, tells you what you can do. [063-MAR, 26, F]*

Participant 034-SAR described their learning about what they needed to work on based on engagement in App A that targeted the whole body:

*Because I wouldn't be strong enough to do it the other way. Yes, and then went down on to the knees and then to try and do the full press-up. But my upper body strength is really, really weak, I found, from doing this. So, but it was great. I think this is certainly something that's been missing from my exercise. I think I have tended to run, if I've done anything, or play netball and a lot of it has been about legs, but not so*

*much upper body. And I think there was a couple of other exercises that I had to go, you know, look for more beyond this app. [034-SAR, 34, F]*

***Preferences of PA influences PA behaviour (Memory, attention and decision processes)***

Participants' decision to use the app was often based on their preferences to engage in PA which, in turn, was influenced by the previous experience. Some of the participants decided not to try the app because of the PA targeted, some opened the apps and disengaged with them, and some enjoyed going back to certain activities they used to do. In addition, the preferences were often expressed as relating to a stable identity around "gym goer", "runner" or "not a natural runner".

The preferential views about PA acted as either enabler or barrier to engagement with the app, e.g., "I'm not someone who is very interested in running" (010-LUC). This preference often meant that participants were not willing to try the PA targeted in the app because of the strong preferences. For example:

*But then I'm not very into this type of thing. I'm not very, doing these things is not my - I'd rather just warm up a little bit, play football. But I'm very football orientated. It's totally different, it is. I'm not into this type of things. [056-CAR, 42, M]*

Below is an example where the participant talked about their issues around running "I'm not a natural runner" (052-WAR) yet they enjoy walking because of the ability to listen to podcasts at the same time:

*I feel there's no pressure. I also, once I'm used to a route, I can do other things which – so I have, I'm, at the moment, for the last year, since Trump was elected, I've been fascinated by American politics. So I'm watching his empire collapse around him and it's extremely*

*interesting (...) And so I listen to that while I'm walking (...)there's no stress, I don't have to have special clothes or, you know, I don't have to move out of people's way or be in people's way." [052-WAR, 44, M]*

### **Apps as potential behaviour change tools**

Below are two extracts from the same person reflecting the good match between their preference for outdoor PA (App B) and incongruence in the case of App A:

*Yes, I love being outdoors, so then that, that ticks that box. You get some fresh air and you get like a physical distance from the other things you're trying to do. Whereas doing that sort of intensive exercises in the same room I'm also doing the studying in, it just gets a bit stressful. [063-MAR, 26, F, App B]*

*And:*

*Maybe it's just not for me, but like I mean people who like doing that kind of intensive muscle-building stuff might enjoy it more. And probably, yes, a way of mixing up the order if it isn't there already. [063-MAR, 26, F, App A]*

There was an aspect to some of the participants about their identity as someone who goes to the gym and that engagement in PA equates to going to the gym. This identity is constructed around various rituals, such as preparation, meeting people in the gym, and the behaviours afterwards, such as the food choices, as illustrated by this participant. As such, none of the apps were able to imitate such experience:

*Yes. There's definitely a weird tragic cult around the gym that is quite embarrassing to be into. I went to a gym in Bali when I was there on holiday which probably a few years ago I would have thought was absolutely mad, who would go to a gym while they're on holiday? But*



*I went and I was absolutely loving it and then the worst part was at the end I went and bought a protein chocolate banana shake and I sat down and drank it. And there was some German guy who was it's the best part of the gym, isn't it? I was yes. [022-MIC, 36, M]*

A preference for a more opportunistic or functional PA was mentioned by participants as opposed to deliberate engagement in structured exercise (as prescribed by both apps). The example of participant 041-AND illustrates well the preference:

*So, rather than working out in my room, I'd rather just explore some part of London. If I lived in a more boring place, then maybe I'd work out more often on my own. [041-AND, 21, M]*

***“Like being reunited with an old friend”: previous PA experience influences engagement in PA (Memory, attention and decision processes)***

Previous experience of PA acted as both, barrier and enabler, depending on the affective experience related to previous PA behaviour. The critical time where PA was mostly conducted throughout participants' lifetimes was during adolescence and possibly whilst attending University where participants would engage in PA and organised sports activities. However, these routines were interrupted when the institution facilitating the engagement was no longer attended. Those who had positive experience with the PA offered in the app seemed to refresh the memory of the positive experience around restarting their PA behaviour.

Stable and the most frequent PA experience was conducted in the contents of institutional activities, either in educational institution or in the context of military service. For example:

*Yes, my experience with running is always been just like running in school, doing PE lessons and I wouldn't say it was a bad experience*

*but it's not something that I find particularly pleasurable. [064-HAY, 33, F]*

However, the experience of this organised PA or sports activities was mixed. Some participants reminisced about positive aspects of engaging in PA with their peers:

*I was more active when I was about between 12 and 15. I was involved with a lot of sport, well, not a lot, but I was – I did football, I played football at lunchtime, I played football with my best mate. Played cricket. I was very, very active and I was really, really active. I was, you do, you do get the endorphins going in you when you've done a bit of exercise. [045-LIA, 26, M]*

Negative experience at school reflected the education system that may lead to negative experience of those that might be perceived as less skilled:

*It's quite hard to get away from this. But those team-picking things were terrible, because I would always be the last or second last to be picked. And people would argue about which team I wouldn't be on rather than the other way round, you know, this kind of thing. So it does nothing for your self-confidence as a kid. [052-WAR, 44, M]*

Often, these activities would stop when the institutional aspect, i.e., extrinsically motivated PA, was removed:

*So, other than that, this really isn't that bad, you could do it anywhere, any time, I mean theoretically I can do it at work too. It's when I used to do when I was in the military. I'd stop every hour and do X number of exercises. Every hour on the hour.*

*(...)*

*But I didn't have – in the military, I had PT, physical training, every day. I had the discipline, I had to stay in shape for the military. And now that I'm out, I don't. And I eat more junk and, yes. [066-CHU, 48, M]*

Indeed, there appeared to be a critical time when PA stopped upon the commencement of adult life with work, time and finance often acting as obstacles, and PA was not on the top priority of tasks to complete:

*Okay, I think it was up and down to be honest. Yes, I mean throughout the years. For example, when I was still at uni, it was okay. And then when I started to do nursing, because of the shifts, every month would be up and down according to my rota. [006-CAT, 26, F]*

### **Apps as potential behaviour change tools**

Previous positive experiences with the running appeared to increase motivation to engage with the apps. For example, in the case of this participant, the running programme brought back certain joy of engaging in PA and acted as an additional source of motivation:

*(...) it is really good to be getting back into running again. Had forgotten how much I enjoy it and will be keeping it up after the end of the trial. Had a spot of shin splint at the end of week one. So had to get out and get a new pair of stability trainers. Got the same model as previously and they were like being reunited with an old friend, made a massive difference.” [054-NIG, 53, M]*

## ***App as a tool providing a structure to fit PA into the routine (Behavioural Regulation)***

### **Apps as potential behaviour change tools**

Participants' attempts to fit PA into their daily routines was a prominent theme. They described how the apps fitted into their schedule by enabling them to plan around their routines. This was especially prominent for App B which provides a nine week structured plan. Because of the provision of the structured plan (as a feature in the app) participants could "visualise" where they will be in x number of weeks. This feature provided motivation to continue the programme:

*It makes you feel like you've actually done something positive with the day, as well as like the studying. I've also got this and then I thought maybe in five weeks, at the end of exams, I might be here somewhere like week six or week seven and be able to have all of those. [063-MAR, 26, F, App B]*

Similarly, App B provides a convenient structure which was appreciated by the participants:

*Yes, especially with the running app because I was like, "Okay, I have one day to rest and then I am going for another run." I don't know, that kept me going. [006-CAT, 26, F, App B]*

The apps also enabled planning PA depending on the time available to engage in PA:

*I'm trying to do, in between, 7 minutes. On the days that I don't run 7 minutes. So, I think they kind of work well. If I've got time. So, my priority will be the running half an hour three times a week and then if I'm not working extra time on those days that I just told you, then I*

*would do the 7 minutes. It's just because if I work from 9 to 10, I'm just knackered. [006-CAT, 26, F, App A]*

Similarly, the prescribed timings of running and walking with the instructions from the audio coach were appreciated and useful:

*It was just good because you, you know, like you get the timer, I like timers and they speak to you and you can actually be outside to use it. And you just press 'play', put it in your pocket and go, like you don't have to watch it all the time. That's what's really frustrating about the first one, like you're constantly like looking to see what to do. Whereas with this one, you're just going. [063-MAR, 26, F, App B]*

### ***Provision of concrete and achievable goals (Behavioural Regulation)***

#### **Apps as potential behaviour change tools**

The app provided goals that were perceived as achievable. In particular, the time specific nature of the PA routines (7 minutes for App A and 30 minutes for 9 weeks for App B) were perceived as motivating for participants, especially in the context of the lack of time to engage in PA as expressed by many participants (and discusses in the theme Lack of time):

*Okay, so, I mean it's good because you don't have an excuse. It's 7 minutes of your time in a day. So, in that sense I think it's very good because you don't have the excuse, 'oh, I don't have time today', because you can, you can put a slot of 7 minutes any time in your day, I think. So, I think, psychological, it's really good in that sense. [006-CAT, 26, F, App A]*

Participant 054-NIG expressed their view about the 30 min sessions with App B:

*It was enough for me and I quite like it. It's a nice time period, it doesn't take too much dedication to find half an hour a day to do it three times a week. And it kind of fitted in with my lifestyle quite nicely. [054-NIG, 53, M, App B]*

By prescribing a specific time-restraint PA session goal to be accomplished, the apps provided a tool to fight the perceived "excuses":

*You have really no excuse not to do it other than yes, I'm lazy. But other than that, you really have no excuse. I was using it every day. [042-MAR, 25, F]*

... and provided a sense of focus:

*A little bit, yes, yes, but I did use the structure of the timing, because that does help, having timed things does help me focus. Yes, that was that one. [063-MAR, 26, F]*

Although the exercises in App A were *high-intensity*, they were short and seemed attainable. For example:

*Because I mean usually, I'd do that, but I just hold it for as long as I like, feel like holding it. But this one it gives you like a time and you're like I know I'm going to fall but it's like 5 seconds left. So, then you like hold it for that extra 5 seconds, I guess. [010-LUC, 19, F, App A]*

#### 6.4.2.3.2 Physical Capability

##### ***Tailoring of the app to the users' starting level of fitness (Physical skills)***

##### **Apps as potential behaviour change tools**

The main reason for the importance of tailoring was so that the PA starting level could be matched to the individual's needs. As discussed in the theme on the importance of emotions in PA (theme 'Emotional influence') a lack of challenge produced feelings of boredom, whereas frustration seemed to follow when the activity was too intense for the participant's level of fitness. This quote reflects well the need for tailoring:

*Yes. But, I don't know, like I think if I was – I think if it had stuff or easier things like a warmup and stuff like a complete beginner – because, to me, that's not a beginner level, because it's too hard to actually do.*  
[063-MAR, 26, F]

Similarly, the "unachievable" nature of the goal that was prescribed by App A cannot be underestimated:

*So, kind of like you can't screen people. 'Do you like being outdoors, do you...?' I don't know. Making it kind of – the main thing for me was that the first thing that you do is achievable and not really too difficult, because that, like – maybe I'm the only person in the world that can't do a press-up. [063-MAR, 26, F]*

Contrary to the previous participant, others explicitly expressed that the level of PA should be assessed and some participants used their own "tailoring" in order to make the PA session more challenging:

*...so on the last run of my second week, the last, really the last run of this app, was I – I think I've switched turns, so I walked when I was*

*supposed to run and the other way round just because I felt like I could run a little bit more and walk a little bit less, rather than the opposite.*  
[028-ADA, 30, F]

Assessment of the starting level PA might also be important because it could make the feedback features more personal, and hence more meaningful:

*Like, to help with progression. So, you can see exactly where you are at stage one. And then, you know, at stage two, this is, you know, how much change that you have had, when you enter in those details afterwards.* [043-JEN, 26, F]

## **6.5 Discussion**

### **6.5.1 Principal findings**

The aim of this study was to explore the acceptability of trial procedures, and to identify and systematically map the enablers and barriers to PA in a sample of physically inactive adults living in urban areas. The findings provide an in-depth understanding around the variability of the PA outcome as observed for the participants in the feasibility trial. The study findings suggest that the trial procedures (e.g., using the apps and wearing accelerometers) were acceptable for the participants. A wide range of influences on PA was identified which had an impact on participants' capability, motivation and opportunity to engage in PA. The motivation to engage with PA using the apps varied substantially and there were environmental factors identified, both physical and social, that seemed to impact on PA. The apps tested in the trial can help with addressing some of the barriers but there are gaps between the obstacles to PA and the app interventions which are not currently addressed. Tailoring of the apps to the users' circumstances (both thinking processes and environmental context) is important.



### **6.5.1.1 Acceptability of the trial design and procedures**

#### **Trial design**

None of the participants had negative views around the trial design. On the contrary, the opportunity to use two apps was seen as beneficial and many participants would compare the apps to consider what PA they might prefer. However, as this was a complex design many of them appreciated the reminders to wear the accelerometer device and the provision of the study schedule. As such, this crossover design will be considered for the use in the definitive trial.

#### **Data collection methods**

Most of the participants did not have any issues with wearing the accelerometer and often reported not being aware of it whilst engaging with their daily tasks. A few participants mentioned the device moving and, specifically, rising up whilst worn. This issue needs to be considered as it might influence the monitoring of PA levels. In addition, the use of EMA was an acceptable data collection method. Many of the participants reported being more self-aware of how they felt when asked to rate their mood using EMA. Further studies should explore the assessment of mood as a potential intervention increasing the awareness of one's feelings and possibly increase the mindfulness of the mood changes following the engagement in PA.

However, there were some issues around the use of the PACO software (which delivered the EMA) that should be considered. Some participants responded positively and some negatively to the barking dog sound of the prompt. Hence in the future trial the sound of the prompt should be tailored to the preference of the individual. A major issue was that some of the participants using Android phone would not receive notifications from the app and a handful of participants received their prompts in the middle of the night. This might render this app unacceptable to use in a definitive trial. Lastly, there were no issues reported with filling in the study questionnaires.

### **6.5.1.2 Enablers and barriers of PA and how apps can help or hinder the PA engagement**

This study adds to the existing evidence on the enablers and barriers of PA behaviour change [366-369]. In addition, it further contributes to the knowledge on how mobile apps can help to increase PA but also demonstrates the limitations of these digital interventions. These influences on PA behaviour and the app-specific PA behaviour are discussed below, facilitated by the use of the COM-B model.

#### **Reflective Motivation**

There were five TDF domains influencing PA behaviour identified within the Reflective Motivation component of the COM-B: Intentions, Beliefs about capability, Goals, Beliefs about consequences, and Behavioural Regulation, which included themes related to the intentions, self-efficacy, health beliefs about the PA and the safety of the app-prescribed PA, conflicting priorities, and app enabling structure to fit PA into daily routines.

Intentions to exercise and self-efficacy are consistent predictors of PA, e.g., [370, 371] and were identified in this study. The apps have increased the awareness of participants' PA which seemed meaningful for some participants. In addition, confidence increased through the ability to follow the apps' PA routines at one's own pace. The sheer engagement in the PA routines seemed to act as an enabler of further engagement in PA through an iterative process of reinforcement.

However, the discrepancy between the intentions of the participants to increase the PA and the reality of the engagement is well documented in research literature [103, 372] and was evident in the context of this study. The trial seemed to have increased the awareness of participants' PA which, although perceived as meaningful to some

participants, did not increase the use of the apps. Whether this awareness was translated into behaviour beyond the intervention period was not assessed in this study.

The pre-conceived health beliefs about the positive and negative consequences of engaging in PA are important and have been shown to influence the decision to conduct PA [373, 374]. In this study, the beliefs about health consequences the exercises targeted by the apps seemed to influence the engagement in PA.

In addition, beliefs about the exercise targeted by the app could sometimes hinder engagement in PA. Specifically, the safety of the PA guidance in the apps, i.e., running in urban spaces (App B) and the lack of in-depth guidance on conducting the exercises and alterations based on PA skills and stamina were perceived by some of the participants as unsafe. It is important that high quality apps targeting PA include safety measures to minimise the potential of injury, and it is specifically important when targeting inactive groups as those untrained are more likely to be injured [375]. Implementing effective injury prevention measures are effective in decreasing the likelihood of injury [376]. However, a review of PA apps' adherence to exercise guidelines [377] showed that apps rarely include such content.

The role and value of deliberate engagement in exercise is difficult with many conflicting life priorities [366, 368] with main concerns being family, friends, work and career progression. In addition, given that the majority of jobs in urban environments are sedentary, participants' prioritisation of work and career progression meant that PA was pushed to a lower rank. Yet, the apps assessed did not include the content feature that would help the user to consider prioritisation exercise to increase the likelihood of engaging in PA.

Those that engaged with the apps seemed to perceive the app tool as a medium that enabled planning and provided a structure that incorporated PA session into their lives. This was most emphasized for the running program (App B) which provides a 9 week plan.

### **Automatic Motivation**

The two TDF domains within the Automatic Motivation component were Emotions and Reinforcement which included the themes relating to emotional influence and provision of achievements (App A).

The negative emotions reported by the participants, such as guilt associated with not engaging in PA, were identified in the responses of the participants and cannot be underestimated. It is unlikely that such negative emotions will lead to behaviour change [378-380]. Anxiety around social commitments and perceived lack of time seemed to influence the lack of engagement in PA. The emotions experienced by the lack of the “match” between the PA and the person, often influenced by previous negative experience, could lead to negative affective state that may lead to further withdrawal from PA.

Positive states following the engagement in PA that was facilitated by the apps produced positive emotions, such as the feelings of being supported, invigorated, uplifted, motivated, “clearing the mind”. This produced a positive feedback and there is evidence that this initial engagement in PA reinforces future engagement [381, 382].

The apps seemed to enable PA by evoking positive emotions through engagement with the PA and the feature of the voice coach (App B). They may hinder the engagement in

behaviour when the PA is perceived as too difficult and when the repetitive nature of the PA is seen as boring.

**“Yes, I really liked it even though I didn’t ever do seven minutes.”**

In the context of emotions, it has to be noted that some participants expressed that they liked the apps but did not engaged in the PA using the apps. This is important and might reflect the gap between the “like” and the “effective” as proposed in the discussion of the Study 2 results (section ‘Why would almost three million users not be able to filter apps that are more likely to be effective?’ Chapter 3)

There is evidence that provision of rewards can be effective in digital interventions [383]. The gamification element seemed to have some impact on some of the participants in this study through reinforcement. This feature (provided by App A) acted as an enabler of PA through providing a sense of achievement and the motivation to collect the rewards.

### **Physical Opportunity**

The TDF domain of Environmental Context and Resources included themes related to: the influence of weather, lack of time, outdoor built environment, and indoor environment.

There is strong evidence that engagement in PA is influenced by weather [275, 276]. Lack of time is one of the most commonly reported barrier to engagement in PA [382, 384], and the built environment, including the aesthetics, neighbourhood safety, and access to green spaces is a frequent determinant of PA [385-387]. These factors were prominent in participants’ responses.

Interestingly, some participants saw the weather as a barrier for doing outdoor activity (targeted in App B), some saw it as a de-motivator to engage in any PA (even that conducted indoors), whereas others did not see the weather as an obstacle. Hence, when a strong and stable motivation is present, the weather was not an obstacle to PA.

One of the advantages of the apps used in this study was their convenience as they provided a prescribed, time-specific PA exercise session (7 min and 30 min) and hence could potentially minimise the issue of time as a barrier. This was not the case for many of the participants with barriers to PA relating to the current life priorities (discussed in the theme concerned with participants' conflicting priorities) and included long working hours, time taken to commute (most often using passive transport system), and family commitments. Yet, those who engaged with PA using the app saw it as an opportunity to fit exercise into their routines.

The built environment, specifically access to green and safe spaces, was important for the participants. App B acted as an enabler through provision of information on how to run safely on the streets (App B). Yet, this information provision was only mentioned by one participant. The findings from the app engagement assessment (section: 'Engagement with the apps' Chapter 5) suggest that participants might not have spent enough time on the apps to be able to engage with the guidance provided in the content of the apps. In addition, although there was an assumption in App A that the user has access to the equipment (however ordinary) to conduct the PA, the nature of living in urban environments meant that some participants had issues with availability of stable chair to stand on, access to bare wall to conduct the wall sit, or even expressed the difficulty conducting jumping jacks because of the consideration for neighbours.

## **Social Opportunity**

The TDF domain of Social Influence included themes related to the social aspects that influence PA engagement.

There is evidence that group-based PA is effective [388, 389] and many users believe that PA behaviour change technology should provide the *human factor* [390]. Many participants described the time when they engaged in PA regularly when they experienced the PA with others. This type of engagement in PA, while exercising, seemed to produce a feeling of enjoyment and entertainment. In addition, it appeared to increase engagement in PA through social commitment.

There was some evidence that the audio coach in App B acted as an enabler when it fulfilled the social role expected by the participants, i.e., being funny, supportive or motivating; this feature may hinder engagement if such expectations were not met. However, neither of the apps included the features of connecting with other users, neither was there a feature of enabling accountability, representing a missed opportunity.

## **Psychological Capability**

The TDF domains within the Psychological Capability component were Knowledge, Skills and Memory, Attention and Decision Processes, and they included the themes relating to the preferences of activities and previous PA experience.

The knowledge and skills around exercise are important factors that play a role in PA engagement [391, 392]. Indeed, in this study misconceptions about PA and the emphasis on the negative aspects, such as the potential of injury seemed to play a role in participant's decision to engage in PA. The content of App B included useful guidance on PA but there was no evidence that many participants read this information. The

evidence in this study suggests that the two apps can act as enablers through the information provision and the positive feedback from the body following a completion of the PA session; apps may hinder PA when the PA targeted by the app may be seen as unsafe or perceived as not appropriate for the participant.

Both the perceived preferences of PA and previous PA experience are important factors in the context of PA behaviour change [371, 381]. In this study, the influence of participants' preferences around certain PA and their *stable identities* (e.g., "not a natural runner") were important. These perceptions seemed to be influenced by previous PA experience including the time during Physical Education or military service. The apps acted as enablers when the preference was congruent with the apps or when the previous experience of the PA was positive; they hindered the engagement when PA did not match the preferences or negative experience connected to the PA was evoked.

### **Physical Capability**

The TDF domain of Physical Skills included the theme related to the tailoring of the app to the user starting level fitness.

Tailoring of the digital interventions is important for engagement [393-395]. In this study, the importance of matching the PA starting level to the individual was evident. The lack of challenge produced the feeling of boredom, and, when the activity was too strenuous for the participant's level of fitness, frustration was experienced. Both states often lead to disengagement with the apps.

### **6.5.2 Strengths**

This study showed that the trial design, the procedures and the app interventions were acceptable to the participants. In addition, it complements the feasibility trial and provides



an in-depth analysis of the real world implementation of two popular publicly available apps.

This study adds to the existing evidence on enablers and barriers to the PA behaviour. In addition, it further contributes to the knowledge how popular PA apps can help to increase PA but also demonstrates the limitations of these digital interventions.

The wide range sample including those that engaged and did not engage with the apps took part in the interviews which provided a comprehensive range of influences on PA behaviour that were included in the findings generated.

This study provided an insight into the experience of using the two popular apps on the market. In addition, not limiting the analysis to the apps, it also shed light on the wider context of PA engagement in physically inactive adults residing in urban environments. As such, the limitations of the apps for facilitating PA were identified and some gaps for improvement were suggested.

The use of TDF and COM-B to map the enablers and barriers of PA and engagement in PA using the apps tested is a strength as the characterisation of the influences of the behaviour can help to identify the potential intervention functions and the BCTs that should be included in interventions to address these influences [107, 108]. The future work is discussed in the main discussion of this thesis (section: 'Immediate future research direction ').

### **6.5.3 Limitations**

During the interviews, one participant revealed they increased their PA levels following their screening questionnaire which potentially means that they would not fulfil the

inclusion criteria. No other participants mentioned increasing their PA but it could not be ruled out.

The findings reflect the participants' salient memory content about the app experience. Data-prompts were used to facilitate recall and the interviews were conducted within a short time following the completion of the trial (not more than a week). However, it is possible that memory bias was present during the interviews.

The influence of biological differences in PA engagement that exist, e.g., sex, ethnicity, age, educational level and BMI, health status [263] were not directly explored in this study.

#### ***6.5.4 Implications for research, practice and policy***

##### **Research**

More research is needed to map the complex interconnected system of engagement in health behaviours in the digital context. Behaviour theory and evidence-based tools, such as the TDF and the COM-B model can help to systematically identify the enablers and barriers to the behaviour in the context of digital interventions. Moreover, the BCW can be used to determine the content of the interventions addressing these specific influences on behaviour.

It is important to conduct more research into how to increase motivation to engage with digital interventions. It is especially important because mere initial engagement with the apps produced immediate positive responses which increased the subsequent engagement with the apps and PA. In addition to the benefits of regular PA, there is also evidence for the immediate positive impact of PA. For example, a single session of moderate-to-vigorous PA can decrease anxiety and insomnia, improve cognition and reduce blood pressure [12].

An important implication for the definitive trial is ensuring the participants feel safe when conducting the PA prescribed by the apps. This might include providing more information about the safe ways to start and continue the PA routine. In addition, although none of the participants reported the use of any other PA apps during the study, such question was not explicitly explored. The process evaluation following the definitive trial should include such question.

Specifically, the safety of the PA guidance in the apps, i.e., running in urban spaces (App B) and the lack of in-depth guidance on conducting the exercises and alterations based on PA skills and stamina were perceived by some of the participants as unsafe. It is important that high quality apps targeting PA include safety measures to minimise the potential of injury, and it is specifically important when targeting inactive groups as those untrained are more likely to be injured [369]. Implementing effective injury prevention measures are effective in decreasing the likelihood of injury [370]. However, a review of PA apps' adherence to exercise guidelines [371] showed that apps rarely include such content. (p. 353).

In the current study, data saturation was reached after 20 interviews. However, there is a scope to use the remaining 16 interviews to explore the differences in those that engaged and did not engaged with the apps. In addition, as there was a substantial variability in the 20% increase in MVPA, it would be of benefit to investigate the differences between those that increased versus decreased their PA. This analysis would potentially enable to categorise the participants into those that are likely to respond to the app interventions. However, due to time and resource constraints, however, the analysis of the differences in responses to PA apps was not conducted.

It is important to establish research evidence on when PA apps can be effective and what their limitations are. This includes identifying the most critical time when these interventions could be introduced.

Digital interventions could potentially help to tackle the environmental barriers to PA, however most apps still rely on the users will, i.e., reflective motivation in a form of self-regulation to change their behaviour. Finally, more research around how to embed a digital intervention within the value system of the users, family, work and career, is of importance.

### **Practice: developers**

The developers of digital interventions should explore the design space to include:

- Features connecting with other users so that PA can be conducted with social support in the real world rather than connecting users in the digital space
- Features using the power of social commitment could be considered as a method to increase engagement in PA
- The guidance around the safety of conducting PA should be emphasised to decrease users' concerns around injury. This should be framed in a positive way [396].
- Designing digital products and services leveraging the users' priorities and values is important for engagement

### **Practice: healthcare professionals**

Although there is evidence that apps can help to change behaviour, healthcare professionals should be aware of the limitation of these interventions.

Health behaviour change using digital interventions may help, however, the evidence suggest that a broader context should be explored when suggesting digital products or services. It is unlikely that digital interventions in themselves will increase health behaviours, in this context PA, at scale and for long-term.

## **Policy**

Digitisation of public and health services is high on the agenda, e.g., [397-399]. However, it is important to first establish the evidence base of when, for what population, in what contexts digital intervention can be effective, and equally effective to face-to-face services.

With competing life priorities, health seems to be prioritised lower among life's most pressing matters. A change of social and cultural norms around engagement in healthy behaviour is needed. Accessibility to and availability of environments conducive to engaging in PA, such as green spaces, safety of the neighbourhood, and aesthetics of the environment are important.

An example of an environmental change at scale is Beelines project [400] which involves transforming the city's cycling network. A project initiated by Greater Manchester's 2<sup>nd</sup> ever Cycling & Walking Commissioner (the 1<sup>st</sup> being in London) is now under way to create over 75 miles separated cycle lane and pedestrian crossing. It will be the largest network in the UK. The process involved mapping disrupted cycle networks and creating connections between the routes for continuous and safer experience. This type of environmental initiatives changing the choice architecture have the potential to change the cultural and social norms by normalising cycling, and encourage walking and running around the city.

When considering the potential of digital interventions to increase access to PA, a new initiative by Sports England is to harness open data to create online services which enable a quick and easy booking system for online classes [401]. With £1.5 million of National Lottery of funding to the Open Data Institute. Developing and implementing this digital initiative may alleviate some of the barriers to accessing PA. However, an evaluation should be in place to assess the impact of such population-level intervention.

### **6.5.5 Conclusions**

Overall, this qualitative study provided some important findings about the experience of taking part in a trial assessing two apps, provided an insight into the enablers and barriers of PA as well as shed light on the experiences of fitting the apps into the routines in adult physically inactive population residing in urban environments. The trial procedures and the app interventions were acceptable. Participants' responses suggested that the opportunity to use two apps was seen as beneficial and many participants compared the apps to inform their PA preferences.

A variety of enablers and barriers to PA engagement were identified providing insights for the quantitative findings of the feasibility trial where a great variability of the responses to the interventions was observed. The two publicly available app interventions can address some of the barriers to PA but it is unlikely that PA apps will be sufficient to address the lack of PA on the population level. Multiple level interventions approaches including collaboration of disciplines are needed. Although intentions, self-efficacy and beliefs are important for health behaviour engagement, incorporating users' value systems as a motivator and targeting environmental factors (both social and physical) facilitating PA engagement are needed in order to change the cultural and social norms around PA.

## **CHAPTER 7. Discussion**

### **7.1 Chapter overview**

In this final chapter of the thesis I provide a summary of the main findings of the thesis and discuss these in relation to the overall aim which was to investigate the public health potential of PA apps. I present my argument that the apps on the market might be effective for certain users, mainly those highly motivated to change. I then discuss the gaps that are unaddressed by the publicly available PA apps and suggest the potential avenues for exploration. I then introduce the perspective of critical health in order to challenge the work in this thesis. I emphasize the need for multidisciplinary approach to digital health and behaviour change. Finally, I present the implications of this research work, and finish with outlining the immediate future research plans.

### **7.2 Principal findings of the thesis**

This thesis investigated the public health potential of publicly available PA apps through the investigation of the app market (Part 1 of this thesis) and through evaluation of the selected apps with inactive users (Part 2). The quality assessment of highly ranked apps showed that, although the popular apps had high usability, i.e. they were predominantly easy to use, there were major issues around the privacy and security of users' data. The restricted inclusion of the behaviour change theory in the app content is also their limitation. Specifically, they mostly target self-regulation processes which rely on deliberate, reflective decision-making of the users to engage with the PA using the apps. In addition, popularity was not associated with likely efficacy providing evidence that popularity does not assure high quality.

The feasibility trial of two popular apps and the subsequent qualitative study showed that such trial is feasible to conduct and acceptable to participants. In addition, the apps assessed have the potential to be effective. Although the trial was not powered to detect a difference, an increase in self-reported PA and in the mediators of PA was observed. In addition, objectively measured PA (20% increase in MVPA) showed that the effects of the apps varied substantially. The opportunity to use two apps was seen as an advantage to the participants. The influences on PA behaviour were identified shedding light on the wider contextual issues of those who are classed as physically inactive and living in urban spaces. Motivation to engage with the apps and to conduct PA seemed to be an overarching issue and this evidence mirrors the research literature on engagement with behaviour change technology [223, 402, 403]. The several environmental and social factors that were identified suggest that, although apps can help to increase PA, their scope may be limited. A multilevel approach is needed to change the activity levels in the population.

### **7.3 What is the public health potential of PA apps on the market?**

In the following sections of this Chapter I consider the implication of the findings in the light of the overall aim of the thesis. I also place this research work within the context of wider research, and I discuss the further research plans.

#### ***7.3.1 The apps on the market are developed for a motivated user***

This thesis proposes that the apps on the market are appropriate for highly motivated users. The top ranked apps focus on providing monitoring and feedback, and enable goal setting and planning. In order to engage with such features, the user needs to engage in an effortful deliberate decision-making processes. However, the prevalence of self-regulation app features is not matched with the UK's population's barriers to PA. For example, as shown in this qualitative study and in the wider literature [366, 368], PA is not on top of the priorities of the potential users and engagement of PA is often based



on past experience. Monitoring, feedback and goal setting might not target such barriers to PA. Hence, the apps may be engaging for a highly motivated user with stable identity characterised by perseverance. Such user will overcome the environmental obstacles to engage in PA. However, when considering the public health potential of those apps, this type of user profile might not be the one targeted by public health interventions.

Indeed, overwhelming evidence shows that human decision-making is often based on urges, emotions and habits. This may often lead to irrational impulsive decisions rather than balanced thought-through decision-making. For example, PRIME theory has advanced our understanding and provided a theoretical structure of human motivation [404]. According to the theory, the human motivational system includes five interacting sub-systems which include plans, evaluations, motives, impulses/inhibitions and response execution. Important influences on the system are: drives, emotional states, arousal, identity and self-control. The theory highlights that motivation is not a stable trait but a dynamic, complex, and volatile phenomenon that changes from moment to moment based on the input from both internal and external environment. As argued by the PRIME theory, reflective decision-making is only part of the process that influences motivation [404, 405].

### **Acknowledging irrationality**

The theories that acknowledged the limitation of human rationality are referred to as dual processing theories [406, 407]. These theories developed simultaneously in various fields of psychology including learning, decision-making, and self-control [408-410]. They distinguish between automatic processing, i.e., habitual and fast which includes impulses/habits/emotional reactions, and controlled, reflective processing which is more effortful and slower, and involves active thought process. The main premise of the dual

process is that engaging the reflective system to override the automatic processes requires cognitive effort, and hence it is limited.

Further evidence for irrationality in human decision-making comes from behavioural economics, a discipline that investigates the systematic deviations from rationality. Standard economics is based on the model of a rational consumer where intentions are formed, pros and cons are weighed up, and optimal decisions are made. In 2002, Daniel Kahneman won a Nobel Prize for his contribution to the psychology of decision-making for his work on the heuristic and cognitive biases. Human decision-making, and the subsequent behaviour, is influenced by a set of heuristics, i.e. mental shortcuts that simplify decision-making and biases, i.e., predictable deviations from rational decision-making [411]. For example, time inconsistency (or present bias) relates to the observation that humans, when making decision, value immediate gain. Hence, the bias can explain the difficulty in placing the emphasis on the long-term detrimental effects of physical inactivity. The impact of heuristics on exercise behaviour has been recently been emphasized by Zenko, Ekkekakis [412].

Williams and Bohlen [413] argued that one of the most relevant heuristics for PA behaviour may be *affect heuristic* which postulates that people are more inclined to engage in behaviours linked to pleasure and avoid those linked to displeasure and pain, and this process is often based on previous experience. It is the fundamental principle of psychological hedonism.

Similarly, Dolan, Kavetsos [414] argued that a shift is needed from educating people about health benefits of PA to methods of making PA perceived as pleasant and enjoyable, which is an “enormous challenge”:

*“Policy can step in to boost exercise participation rates by promoting it as a fun, pleasurable, activity. Happiness is part of the feedback individuals derive from exercising, which can be viewed as an important mechanism encouraging further engagement in that behaviour” [414, p. 1372]*

In summary, there is vast literature on the importance of going beyond the rational decision-making processes that predominantly characterise the main functionality of the PA apps on the market. Specifically, PRIME theory explains the complexity of the dynamic motivational system and stresses the importance of influencing moment-to-moment motivation to increase the likelihood of engaging in PA behaviour. Both, dual process theories and behavioural economics emphasize the impact of the automatic process involved in decision-making, and the importance on focusing of the enjoyable aspects of movement. This Discussion Chapter will now move to recommend how the automatic processes can be targeted.

### **7.3.2 Going beyond rational decision-making – Intervention design**

The *nudge* approach to intervention development aims to change peoples’ choices and behaviours utilising the biases, routines and habits described above. An example of a nudging in the context of PA is to making stairs more attractive and prominent or making cycling more viable through cycle hiring scheme. A type of intervention that aims to change the environment to facilitate the healthier choice is referred to as *choice architecture*. Both nudge and choice architecture have been used to increase health behaviours including in PA [415-417].

Recent research work aimed to develop a system of classifying interventions that alter micro-environments, i.e., the proximal physical environment. The result is the typology of interventions in proximal physical micro-environments (TIPPME) [418]. Although the

framework has been developed based on the research into the purchase and consumption of tobacco, alcohol and food, it has recently been applied to PA. A recent review of 35 PA interventions using nudge utilised TIPPME found that most of the intervention (n= 26) aimed to increase stair use, used exercise commitment (n=4), general PA (n=4) and walking (n=1). All studies focused on micro-level, i.e., individual-level behaviour, and they identified clear opportunities for digital interventions delivered in a pragmatic, real world setting as these were rarely used in the studies identified [419].

### ***7.3.3 Technology can have a place in creating social connections***

In addition to physical environment, the influencing power of social environment should be utilised to affect behaviour change. The prevalence of loneliness continues to increase as the research evidence on the detrimental impact of loneliness also accumulates [420, 421]. The sense of belonging protects against loneliness and participation in groups are considered the active component of many therapies. Although group interventions need to be sensitive to ethno-religious identities [422], harnessing the power of groups can facilitate health behaviour change. For example, social support is an effective methods to increase goal attainment [423, 424]. The influencing social factors on PA were also prominent in participant interviews in this thesis. Yet, the content analysis of the apps (Study 1) showed that most of the features that target social support in popular PA apps do not go beyond the option of sharing users' PA session results on social media, and some apps provide access to in-app communities with similar functions of sharing users' progress.

Although digital technology can emphasize the individualistic nature of a Western culture [425, 426], yet, it can also have the potential to affect behaviour change if technology is used as a medium to connect people with similar interest in the real world [427]. For example, Footy Addicts is a web and app service that enables the users to join football

in local areas. Combining PA with socialising might help to decrease the levels of low mood, create connections within and between the communities, and, in the long-term, help to prevent diseases.

#### ***7.3.4 Away from the dichotomous question***

There is evidence that digitally delivered interventions can have an effect on PA engagement, e.g., [84, 129, 428-437]. The question that remains is: what are the critical moment in the behaviour change journey when digital technology is most likely to act as a propeller to change? As found in this thesis, there was a great variability of the effect of the two apps. However, some participants increased their PA and there was evidence of the beneficial effect of the apps in participants' recollection of their experiences. Still, it is unknown when are the most critical times when the apps should be implemented. It is most likely that tailoring of the digital technology to the user's circumstances would be most beneficial.

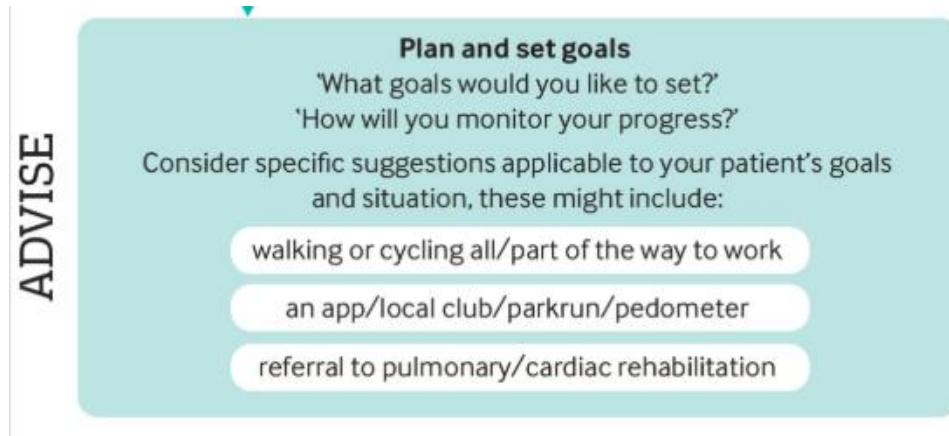
#### ***7.3.5 The importance of tailoring***

Indeed, tailoring of the digital health tools to user preferences is paramount as manifested in the qualitative interviews of this thesis, and wider literature [223, 427]. A short set of questions could be used to gauge user preferences and fitness level and then to tailor the digital interventions to their needs and likes. An example of such tailoring is the Good Thinking service which provides access to digital tools targeting mental health. Their short self-assessment questionnaire is used to tailor the digital intervention displayed to the user based on the result of the survey. This set of questions is the most used feature of the website [438]. Based on the proposed citizen science methods (discussed further), it will be possible to create a more sophisticated profiling of the users so that digital interventions could be matched to their needs and so increase long-term engagement and increase the likelihood of effectiveness.

A promising field of responsive tailoring is Just in Time Adaptive Intervention (JITAI) [439]. JITAIs may be particularly suited for developing PA interventions that utilise time-varying information (e.g., the user's progress towards the goal) to implement dynamic strategies to support behaviour change [440]. Indeed, a recent review of JITAIs found that common BCTs used in these interventions include self-regulation techniques, such as goal setting, feedback, and action planning. The authors concluded that it is crucial to underpin the development of JITAIs in theory [441]. As presented in this thesis, behaviour science tools, such as the COM-B model and TDF may be used to assess the influences on PA behaviour. In addition, appropriate BCTs may be used to develop the JITAI to target these influences [113].

Haseler, Crooke [384] proposed an approach to discuss PA in a primary care consultations, called "ask-assess-advise". The advice part relating to behaviour change is presented in Figure 56. At the time of writing this thesis, there was no standardised questionnaire to assess the patient' characteristics in order to tailor the PA intervention. In the future, based on big data from quantitative PA assessment as well as subjective experiences, it may be possible to develop a standardised questionnaire to assess when people are most open to receiving the advice to increase their PA. By suggesting small, incremental and achievable goals the clinicians could impact the decision to start the process of change. In addition, a tailored digital product/service could be suggested based on the user preferences, level of motivation and the environmental circumstances. Lastly, an alternative to digital interventions could be offered if such a mode of delivery would not be appropriate.

Figure 56: The *advice* stage relating to behaviour change in the ‘ask-assess-advise’ approach (reproduced with permission from [384])



### **7.3.6 Citizen science: the what, for whom, in what context**

There still remains the question of WHICH PA apps have the most potential to change behaviour. However, RCTs may not be appropriate to compare the multitude of apps on the market, neither they are financially feasible. Alternatives to RCTs, as evidenced in this thesis, can include the use of crossover trial. Citizen science might be a more appropriate data collection method to answer the question of what, for whom, under what circumstances may be effective for behaviour change to inform which apps might have the most potential to engage the participants. An example of the use of citizen science is myCircadianClock which is an app that collects daily lifestyle data that can be used by the user whilst contributing to research on the effects of circadian rhythm on health [442]. Based on the data gathered through the involvement of the users, it might be possible to tailor the digital interventions to the profile and the circumstances of the users to shed more light on the holy grail of digital technology: what, for whom, and under what circumstances.

### **7.3.7 Critical approach**

I recognised that it was important to discuss the wider issues around *behaviour change and digital technology* when considering the potential of apps to increase PA on a population level. Hence, I brought the critical health perspective, i.e., the study of the influence of the socio-structural processes and power relations on health and illness [443], into the discussion. Introducing this critical lens was important for two main reasons: 1) to reflect on the discipline I associated myself with – health psychology, 2) to place my work in this thesis in the wider public health system.

#### **7.3.7.1 Digitised healthy citizen**

Within the context of growing digitisation of health and healthcare, Lupton [68] takes a step back and provides a critical perspective around public health messages around self-care and self-management, i.e., encouraging active managing of citizens' health. The language often used by developers of digital health emphasizes the choice and control of the consumer to make *right* decision about their health. Lupton coined the term “digitally engaged patient” [444] where, within the digital context, patients are encouraged to use digital technology to seek information but also to self-monitor, engage in health behaviours to prevent disease, and self-manage their health conditions. She argues that the danger of this notion is that it shifts the responsibility on the individual rather than the state [445, 446]. Such messages, for example, through public health campaigns, understate other ways of caring for citizens, such as collective and state-supported initiatives [447].

This criticism reflects the discipline of health psychology, and, specifically, the use of behaviour change models, in particular social and cognitive behaviour change model which focus on the individualistic approach. These models take for granted the assumption about the need to persuade people to change their behaviour and adopt



practices that are seen as "right". As such, mainstream health psychology is ignoring the environmental dimensions of digital health. In such a way, health psychology encourages the focus on people to take responsibility to increase their health and detracts from the socio-cultural context in which health practices are performed whilst ignoring the lived experience [448, 449]. It has to be noted that this approach is not limited to neither digital research nor health psychology but reflects the trends in psychology as a discipline [68]. Although the COM-B model used in this thesis acknowledges the environmental factors involved in behaviour change, it is important to explicitly acknowledge the issues around health behaviour when targeting individual behaviour change. Hence, it is important, in the context of this thesis, to place PA apps as one of the tools that can facilitate engagement in PA within the wider social, cultural and political environment.

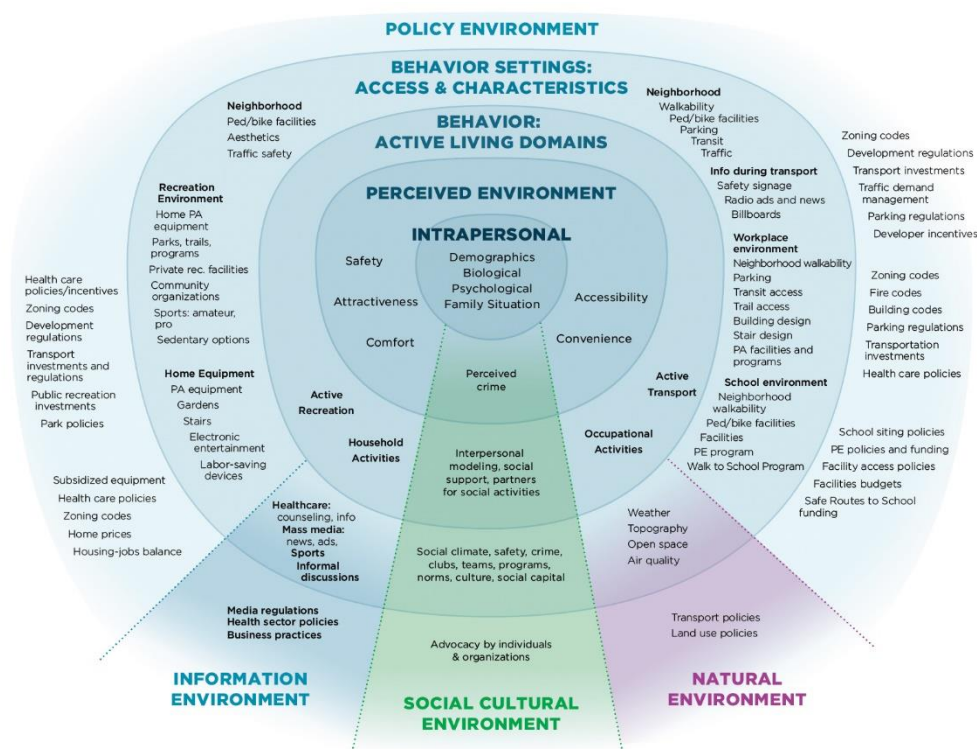
### **7.3.8 The need for multisector and whole system approach**

Ultimately, it will take both internal (person-based) and external (environment-based) changes to affect health and well-being on a population level. Indeed, PA is increasingly being defined as a major, complex, and multisector issue [450]. The analysis of PA policies in England argue that there is a need for multi-level, cross-sectoral investments that integrate public health campaigns with infrastructure to increase the opportunities in the environment to be more active [87].

Shifting away from PA and moving into *active living* has been suggested and might be more proportionate to the overarching goal of reducing the burden to the NHS through the prevention of noncommunicable diseases [451-453]. For example, in line with Lupton's critical insights, the social ecological model recognises the complex intertwined relationships between the individual, their environment and the policies that affect both. Individuals may be responsible for behaviour change to some degree; however, they are influenced by their social, physical and political environments. A social ecological model, with proposed four domains of active living including active recreation and transport,

and occupational and household activities, is presented in Figure 57 [454]. This visualisation illustrates well the complexity of the PA engagement. It highlights that social and cultural change is needed to embed PA within individuals' values and life priorities, i.e., as part of a lifestyle rather than *detached* behaviour.

Figure 57: Social ecological model of active living (reproduced with permission from [454])



Lastly, risk of disease is unlikely to be explained by a single behaviour. For example, a complex system of behaviour and behavioural outcomes including diet, PA and body weight are more likely to explain risk of cancer [455]. In addition, even in sufficiently active individuals, daily sedentary time is an independent risk factor for all cause mortality, cardiovascular disease, and diabetes [456].

### **7.3.9 *The need for a multidisciplinary work: shedding the disciplinary self-identity***

Everyday life, where health and behaviour is experienced, is not made up of medicine, psychology, technology, etc. These seemingly independent areas of knowledge/skills are organised into disciplines to advance the different aspects of understanding. However, translating and implementing the learning into practice needs consolidation of these entities into *real life*. However, interventions are often developed (and funded) focusing on specific, often singular, behaviours from a single discipline point of view. There is a need to reduce the separation by bringing together various disciplines and approach the intervention development as a system taking into account the “whole” being, the social and built environment, and the powerful and dynamic influence of automatic (unconscious) processes on behaviour. Targeting a single behaviour from a single discipline is unlikely to bring about the change needed to save the NHS.

## **7.4 Immediate future research direction**

This section outlines the proposed next step following the completion of this research work.

### **7.4.1 *Definitive trial***

Part 2 of this PhD, which focused on the assessment of the two popular PA apps, demonstrated that this crossover trial is both feasible to conduct, and acceptable to participants. Hence, based on the lessons learnt from running the feasibility trial, the next steps will be to conduct a definitive trial of the two apps. The opportunity to use two apps was seen as an advantage to participants. Participants compared these two interventions to each other forming preferences of what they would and would not like to use. This seemed to mimic the market where apps are downloaded, tried, and discarded if

perceived as not useful. Hence, the lessons learned from conducting this feasibility trial will be used to refine the development of the protocol for the definitive trial.

In comparison to this feasibility study, the definitive crossover trial will include longer follow-up periods to assess the behavioural effects and the maintenance of PA using the apps. Accelerometer device will be used to measure the primary outcome because it is a gold standard in PA research [337, 338]. A continuous variable will be used as it is more powerful to assess the real difference in PA. However, the process evaluation will compare those participants that respond to the app interventions to those that do not, as variability in PA outcome is expected, as learned in this feasibility study.

In addition to the self-reported engagement scale used in this study, it would be of benefit to assess the objective measures of usage data. This will mean that a close collaboration between the developers and the researchers is needed to access to the usage data. Moreover, as this qualitative study showed, participants had issues with the safety of the PA prescribed in the apps, additional information around injury prevention will be included.

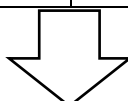
The advantage of conducting a definitive crossover trial includes the ability to assess two apps to increase PA in inactive population. Fewer participants are needed for a crossover trial in comparison to other designs where each participant only receives a single intervention. However, a disadvantage of this type of design is that it treats the app intervention as whole and hence individual features or BCTs are not assessed. As participant will be randomised to the app use, it will not closely represent a real-world situation. However, this is a common issue in experimental studies.

The preliminary findings showed clear differences in self-reported measures, both psychological and behavioural. The variability in the 20% increase in MVPA using accelerometer data suggest it might be important to investigate who increased and decreased their PA so that the apps could be tailored to the individuals responsive to the intervention. Hence, there is a scope for investigating the differences in those participants that decreased and increased their MVPA. This could be done using qualitative exploration of the data gathered as part of this PhD. These findings could inform the development of a questionnaire to assess the readiness to use a digital aid to increase PA.

#### **7.4.2 Strategic behavioural analysis**

Study 1 helped to answer the question around what the apps *do* through the characterisation of the BCTs and the feature extraction. Further, the identification of enablers and barriers to PA (Study 4) helped to answer the question about what *should* the apps address. Table 68 displays how the studies in this thesis complement each other. In the next section I present how I would like to take this research learning further by conducting a strategic behavioural analysis.

Table 68: A diagram of complimentary studies of this PhD and the next step proposed

Study	Question answered in this PhD	The output of the method used
Study 1	What is in the apps now?	Content of apps: BCTs, app features
Study 4	What should the apps target?	Identified enabler and barriers to PA
		
Further work	What are the gaps and opportunities for improvement?	Strategic behavioural analysis

## **Theoretical congruence between the influences on the PA behaviour and the content of the popular PA apps**

By using a systematically developed tool, it is possible to identify the intervention functions, policy categories, and BCTs that would be most likely to target the underlying influences on the behaviour. This process is termed as behavioural analysis which is a vital stage of theoretically-driven intervention development [457, 458]. This study shed light on the influences of PA behaviour in the sample of participants that took part in the trial (Study 4), however a high quality behavioural analysis will involve synthesising the evidence base on the key drivers of behaviour from a wider research literature.

The mapping of the influences on the PA behaviour onto the TDF domain in this study provides the ground for future work. The BCW matrices (presented in Appendix Q) can be used to ascertain the most relevant intervention functions and policy categories to be targeted to influence the enablers and barriers [113, 114]. Most relevant to this thesis, however, is a recent work based on a literature review of 277 studies, a consensus of 105 experts in behaviour change and a triangulation of the two, produced an evidence-based matrix mapping TDF to BCT Taxonomy (v1) [459]. For example, The TDF: Environmental Context and Resources was a prominent influence on the PA behaviour as found in the qualitative interviews (Study 4). This domain is most likely to be targeted by BCTs:

- 3.2. Social support (practical),
- 7.1. Prompts/cues,
- 7.5. Remove aversive stimulus,
- 12.1. Restructuring the physical environment,
- 12.2. Restructuring the social environment,
- 12.3. Avoidance/reducing exposure to cues for the behaviour,

- 12.5. Adding objects to the environment.

Although 3.2. *Social support (practical)*, and 7.1. *Prompts/cues* were identified in some of the apps, none of the other BCTs were identified in the review and content analysis (Study 1). Hence, the TDF mapping of the enablers and barriers showed that there is some overlap between the influences on behaviour and what is targeted in the apps. However, there are many incongruences.

The process of exploring the coherence between the influences on the behaviour (in this context PA engagement) and the interventions (in this context existing PA apps) enables to identify the missed opportunities for intervention content, and is termed *strategic behavioural analysis*. The COM-B and TDF has recently been applied in an increasing number of systematic reviews as a framework for synthesising behavioural determinants across qualitative and quantitative studies reporting perceived enablers and facilitators of a range of behaviours [460-462].

Hence, the my proposed next steps will be to conduct a strategic behavioural analysis to explore the influences on the behaviour and to recommend interventions that would be most appropriate to target these influences. This work would be alongside the trial assessing the apps. The rationale for this parallel work is to assess the effectiveness of apps (following the feasibility trial) but also to address the broader issues of engagement in PA, as discussed in this Chapter. As such, the question of interest would shift from: 'Are apps effective?' to 'When are apps effective?' i.e., at what stage the apps could be introduced to increase the likelihood of engaging in PA.

### **7.4.3 Strengths**

This thesis has several strengths. First, by focusing on the app market, i.e., the interventions (Part 1) as well as on the potential user, i.e., the target users (Part 2), this thesis gained a comprehensive understanding of both the digital health technology and behaviour change in the context of physical inactivity. Second, the methodology used in this thesis included a range of qualitative and quantitative methods. The use of novel research methods (e.g., content analysis of highly-ranked PA apps, crossover trial to assess apps, EMA, data-prompted interviews) enabled rich data collection and analysis. Third, the insights from this thesis are set in the *real world*. The evaluation of the top-ranked apps enabled this thesis to focus in the digital health technology that is, at least, used once by the public as the download rates for the top-ranked apps are in millions. Fourth, the methodology is pragmatic and applied to health psychology in public health. As such, the quality framework developed in Study 1 can be used to assess the quality of technology in other behaviours and using various modes of digital delivery, for example online interventions. In addition, the novel application of a crossover trial to assess the two digital interventions can be applied to other behaviours and digital health interventions.

### **7.4.4 Limitations**

First, the thesis had a broad scope including the assessment of the state-of-the-art PA apps on the market, and the potential effects of the apps on the users. As such, the review and content analysis was conducted on top-ranked 65 PA apps from a pool of 121. Although it is unknown how many PA apps are on the market, the number of general health and wellness apps exceeds 325, 000 [81]. In addition, it was possible to assess only two of the most popular apps in a trial. Second, the inclusion of BCTs was used to represent effectiveness; however, this variable does not substitute experimental studies, such as RCTs. Third, the generalisability of the findings from the crossover trial and the



qualitative study is limited to the UK population living in urban environments. Fourth, the short periods of app assessment were used to assess feasibility and acceptability. For the definitive trial, longer intervention period is needed to assess the behavioural effect of the apps. Fifth, as I did not build the apps assessed, I had no control over the content and functionality of the apps. Sixth, this thesis assessed the quality of the apps on the market. However, due to nature and the knowledge base around health apps, it was not possible to ascertain how many BCTs are sufficient or optimal for behaviour change. Hence, making a recommendation using firm cut-off points to distinguish *good* from *bad* apps was not possible. Lastly, little knowledge was generated around when the PA apps can be most effective, i.e., what specific user characteristics are matched to app intervention features and content to produce the optimal effect.

#### **7.4.5 Concluding remarks**

Physical inactivity contributes to as many deaths in the UK as smoking [49]. Advancing health technology and prevention are two of the three main priorities of the current Health and Social Care Secretary. This thesis aimed to investigate the potential of current publicly available apps to increase PA on the population level by assessing the quality of the apps on the market and evaluating the potential of the apps with inactive users. This was achieved through the use of a range of qualitative and quantitative methods. The quality framework assessing safety, effectiveness and user experience was developed and applied to a sample of highly ranked PA apps. No relationship was found between popularity and likely efficacy suggesting that popularity does not assure high quality, and what is liked may not be what is likely to be effective. The evaluation of two highly ranked apps showed that the study is feasible and acceptable. The impact of the PA app interventions showed promising results with an impact observed for self-reported PA, intentions and exercise self-efficacy. There was a range of influences on PA behaviour identified including the possible use of apps as assisting in PA engagement. The conclusion of this thesis is that PA app interventions have their place in the wider

healthcare system, and the further research will focus on when, where and at what stage in the behavioural journey they should be introduced to increase the likelihood of realising their potential to support the user with sustainable behaviour change.

## References

1. Pedersen, B.K. and B. Saltin, *Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases*. Scand J Med Sci Sports, 2015. **25 Suppl 3**: p. 1-72.
2. Knols, R., et al., *Physical exercise in cancer patients during and after medical treatment: a systematic review of randomized and controlled clinical trials*. J Clin Oncol, 2005. **23**(16): p. 3830-42.
3. Booth, F.W., C.K. Roberts, and M.J. Laye, *Lack of exercise is a major cause of chronic diseases*. Compr Physiol, 2012. **2**(2): p. 1143-211.
4. ukactive, *Turning the tide of inactivity*. London: UK Active. Retrieved, 2014. **1**(9): p. 15.
5. Chen, J.J. and Y. Lee, *Physical activity for health: evidence, theory, and practice*. J Prev Med Public Health, 2013. **46 Suppl 1**: p. S1-2.
6. Zschucke, E., K. Gaudlitz, and A. Strohle, *Exercise and physical activity in mental disorders: clinical and experimental evidence*. J Prev Med Public Health, 2013. **46 Suppl 1**(Suppl 1): p. S12-21.
7. UK Chief Medical Officers, *UK Chief Medical Officers' Physical Activity Guidelines*. 2019.
8. Piercy, K.L., et al., *The physical activity guidelines for Americans*. Jama, 2018. **320**(19): p. 2020-2028.
9. Caspersen, C.J., K.E. Powell, and G.M. Christenson, *Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research*. Public Health Rep, 1985. **100**(2): p. 126-31.
10. Khan, K.M., et al., *Sport and exercise as contributors to the health of nations*. The Lancet, 2012. **380**(9836): p. 59-64.
11. Ainsworth, B.E., et al., *2011 Compendium of Physical Activities: a second update of codes and MET values*. Medicine & science in sports & exercise, 2011. **43**(8): p. 1575-1581.
12. Health, U.D.o. and H. Services, *2018 Physical activity guidelines advisory committee scientific report*. 2018.
13. Murphy, M.H., S.N. Blair, and E.M. Murtagh, *Accumulated versus continuous exercise for health benefit*. Sports medicine, 2009. **39**(1): p. 29-43.
14. Oberbach, A., et al., *Effect of a 4 week physical training program on plasma concentrations of inflammatory markers in patients with abnormal glucose tolerance*. European Journal of Endocrinology, 2006. **154**(4): p. 577-585.
15. Pitsavos, C., et al., *The associations between physical activity, inflammation, and coagulation markers, in people with metabolic syndrome: the ATTICA study*. European Journal of Cardiovascular Prevention & Rehabilitation, 2005. **12**(2): p. 151-158.
16. Silverman, M.N. and P.A. Deuster, *Biological mechanisms underlying the role of physical fitness in health and resilience*. Interface focus, 2014. **4**(5): p. 20140040.
17. Ertek, S. and A. Cicero, *Impact of physical activity on inflammation: effects on cardiovascular disease risk and other inflammatory conditions*. Archives of medical science: AMS, 2012. **8**(5): p. 794.
18. Nader, G.A. and I.E. Lundberg, *Exercise as an anti-inflammatory intervention to combat inflammatory diseases of muscle*. Current opinion in rheumatology, 2009. **21**(6): p. 599-603.
19. Metsios, G.S., et al., *Vascular function and inflammation in rheumatoid arthritis: the role of physical activity*. The open cardiovascular medicine journal, 2010. **4**: p. 89.

20. Waschki, B., et al., *Physical activity is the strongest predictor of all-cause mortality in patients with COPD: a prospective cohort study*. Chest, 2011. **140**(2): p. 331-342.
21. Onur, E., et al., *The beneficial effects of physical exercise on antioxidant status in asthmatic children*. Allergologia et immunopathologia, 2011. **39**(2): p. 90-95.
22. Giovannucci, E., *Metabolic syndrome, hyperinsulinemia, and colon cancer: a review*. The American journal of clinical nutrition, 2007. **86**(3): p. 836S-842S.
23. Xue, F. and K.B. Michels, *Diabetes, metabolic syndrome, and breast cancer: a review of the current evidence*. The American journal of clinical nutrition, 2007. **86**(3): p. 823S-835S.
24. Cotman, C.W., N.C. Berchtold, and L.-A. Christie, *Exercise builds brain health: key roles of growth factor cascades and inflammation*. Trends in neurosciences, 2007. **30**(9): p. 464-472.
25. Blumenthal, J.A., et al., *Effects of exercise training on older patients with major depression*. Archives of internal medicine, 1999. **159**(19): p. 2349-2356.
26. Nabkasorn, C., et al., *Effects of physical exercise on depression, neuroendocrine stress hormones and physiological fitness in adolescent females with depressive symptoms*. European journal of public health, 2006. **16**(2): p. 179-184.
27. Activity, I.S.f.P. and Health, *The Bangkok declaration on physical activity for global health and sustainable development*. Br J Sports Med, 2017. **51**(19): p. 1389-1391.
28. Sallis, J.F., et al., *Progress in physical activity over the Olympic quadrennium*. The Lancet, 2016. **388**(10051): p. 1325-1336.
29. Deng, T., et al., *Obesity, inflammation, and cancer*. Annual Review of Pathology: Mechanisms of Disease, 2016. **11**: p. 421-449.
30. DoHaHS, D., *Physical Activity Guidelines Advisory Committee report, 2008. To the secretary of health and human services. Part A: executive summary*. Nutr Rev, 2009. **67**: p. 114-120.
31. Cannioto, R., et al., *LIFETIME PHYSICAL INACTIVITY IS ASSOCIATED WITH LUNG CANCER RISK AND MORTALITY*. Cancer treatment and research communications, 2018. **14**: p. 37-45.
32. Islami, F., et al., *Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States*. CA: a cancer journal for clinicians, 2018. **68**(1): p. 31-54.
33. Scerpella, J.J., et al., *Increased leg bone mineral density and content during the initial years of college sport*. The Journal of Strength & Conditioning Research, 2018. **32**(4): p. 1123-1130.
34. Watson, S.L., et al., *High-intensity resistance and impact training improves bone mineral density and physical function in postmenopausal women with osteopenia and osteoporosis: the LIFTMOR randomized controlled trial*. Journal of Bone and Mineral Research, 2018. **33**(2): p. 211-220.
35. Chou, C.-H., C.-L. Hwang, and Y.-T. Wu, *Effect of exercise on physical function, daily living activities, and quality of life in the frail older adults: a meta-analysis*. Archives of physical medicine and rehabilitation, 2012. **93**(2): p. 237-244.
36. Warburton, D.E., et al., *The risk-benefit paradox of exercise*. BC Medical Association Journal, 2016. **58**: p. 210-218.
37. McKercher, C., et al., *Physical activity patterns and risk of depression in young adulthood: a 20-year cohort study since childhood*. Social psychiatry and psychiatric epidemiology, 2014. **49**(11): p. 1823-1834.
38. Mura, G., et al., *Exercise as an add-on strategy for the treatment of major depressive disorder: a systematic review*. CNS spectrums, 2014. **19**(6): p. 496-508.

39. Kratz, A.L., D.M. Ehde, and C.H. Bombardier, *Affective mediators of a physical activity intervention for depression in multiple sclerosis*. Rehabilitation psychology, 2014. **59**(1): p. 57.
40. Greenwood, B.N., et al., *The protective effects of voluntary exercise against the behavioral consequences of uncontrollable stress persist despite an increase in anxiety following forced cessation of exercise*. Behavioural brain research, 2012. **233**(2): p. 314-321.
41. Nishijima, T., et al., *Cessation of voluntary wheel running increases anxiety-like behavior and impairs adult hippocampal neurogenesis in mice*. Behavioural brain research, 2013. **245**: p. 34-41.
42. Lynch, W.J., et al., *Exercise as a novel treatment for drug addiction: a neurobiological and stage-dependent hypothesis*. Neuroscience & Biobehavioral Reviews, 2013. **37**(8): p. 1622-1644.
43. Peterson, A.B., D.P. Hivick, and W.J. Lynch, *Dose-dependent effectiveness of wheel running to attenuate cocaine-seeking: impact of sex and estrous cycle in rats*. Psychopharmacology, 2014. **231**(13): p. 2661-2670.
44. Mattson, M.P., *Interventions that improve body and brain bioenergetics for Parkinson's disease risk reduction and therapy*. Journal of Parkinson's disease, 2014. **4**(1): p. 1-13.
45. Bielak, A.A., et al., *Preserved differentiation between physical activity and cognitive performance across young, middle, and older adulthood over 8 years*. Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 2014. **69**(4): p. 523-532.
46. Beier, M., et al., *Improved physical fitness correlates with improved cognition in multiple sclerosis*. Archives of physical medicine and rehabilitation, 2014. **95**(7): p. 1328-1334.
47. Tian, Q., et al., *Physical activity predicts microstructural integrity in memory-related networks in very old adults*. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 2014. **69**(10): p. 1284-1290.
48. Kokkinos, P., *Physical activity, health benefits, and mortality risk*, ISRN *Cardiol.*(2012) 718789. 2012.
49. Public Health England, *Everybody active, every day: a framework to embed physical activity into daily life*. 2014, PHE publications gateway.
50. World Health Organisation. *Physical Inactivity: A Global Public Health Problem*. Available from: [https://www.who.int/dietphysicalactivity/factsheet\\_inactivity/en/](https://www.who.int/dietphysicalactivity/factsheet_inactivity/en/).
51. Yang, L., et al., *Trends in Sedentary Behavior Among the US Population, 2001-2016*. Jama, 2019. **321**(16): p. 1587-1597.
52. World Health Organisation, *Reduce physical inactivity*, in *Noncommunicable diseases: Campaign for action – meeting the NCD targets*. 2016.
53. Sports England, *Tackling inactivity: a guide to sport england's approach and investment*. 2016.
54. Health & Social Care Information Centre. *Statistics on Obesity, Physical Activity and Diet*. 2016 27/2/17]; Available from: <http://content.digital.nhs.uk/catalogue/PUB20562/obes-phys-acti-diet-eng-2016-rep.pdf>.
55. Craig, R., J. Mindell, and V. Hirani, eds. *Health Survey for England: physical activity and fitness*. 2008, The Information Centre: London.
56. Lee, I.-M., et al., *Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy*. The lancet, 2012. **380**(9838): p. 219-229.
57. Arem, H. and M. Irwin, *A review of web-based weight loss interventions in adults*. Obesity reviews, 2011. **12**(5): p. e236-e243.

58. Warburton, D.E. and S.S. Bredin, *Health benefits of physical activity: a systematic review of current systematic reviews*. Current opinion in cardiology, 2017. **32**(5): p. 541-556.
59. Moore, S.C., et al., *Leisure time physical activity of moderate to vigorous intensity and mortality: a large pooled cohort analysis*. PLoS medicine, 2012. **9**(11): p. e1001335.
60. Huckvale, K., et al., *Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment*. BMC Medicine, 2015. **13**(1): p. 1-13.
61. Price, M., et al., *mHealth: a mechanism to deliver more accessible, more effective mental health care*. Clinical psychology & psychotherapy, 2014. **21**(5): p. 427-436.
62. Glasgow, R.E., et al., *Interactive behavior change technology: a partial solution to the competing demands of primary care*. American journal of preventive medicine, 2004. **27**(2): p. 80-87.
63. Schüll, N.D., *Data for life: Wearable technology and the design of self-care*. BioSocieties, 2016. **11**(3): p. 317-333.
64. eBizMBA. *Top 15 most popular health websites February 2017*. 2017; Available from: <http://www.ebizmba.com/articles/news-websites>.
65. Moorhead, S.A., et al., *A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication*. Journal of medical Internet research, 2013. **15**(4): p. e85.
66. Nabi, R.L., A. Prestin, and J. So, *Facebook friends with (health) benefits? Exploring social network site use and perceptions of social support, stress, and well-being*. Cyberpsychology, Behavior, and Social Networking, 2013. **16**(10): p. 721-727.
67. Bender, J.L., M.-C. Jimenez-Marroquin, and A.R. Jadad, *Seeking support on facebook: a content analysis of breast cancer groups*. Journal of medical Internet research, 2011. **13**(1): p. e16.
68. Lupton, D., *Digital health: critical and cross-disciplinary perspectives*. 2017: Routledge.
69. Van Dijk, J. and K. Hacker, *The digital divide as a complex and dynamic phenomenon*. The information society, 2003. **19**(4): p. 315-326.
70. Ofcom, *UK Communications Market Report 2018*. 2018.
71. Digital Exclusion Heatmap. 2017; Available from: <http://heatmap.thetechpartnership.com/>.
72. Bank, L., *UK Consumer Digital Index 2018*, in Lloyds Bank, London. 2018.
73. Anderson, M., *Technology device ownership, 2015*. 2015: Pew Research Center.
74. Poushter, J., *Smartphone ownership and internet usage continues to climb in emerging economies*. Pew Research Center, 2016. **22**: p. 1-44.
75. Koetsier, J. *Battle of the mobile sexes: women install 40% more apps, spend 87% more than men*. [cited 2019 23/11/19]; Available from: <https://venturebeat.com/2013/04/26/battle-of-the-mobile-sexes-women-install-40-more-apps-spend-87-more-than-men/>.
76. Nielsen. *Hacking health: how consumers use smartphones and wearable tech to track their health*. 2014 [cited 2019 23/11/19]; Available from: <https://www.nielsen.com/us/en/insights/article/2014/hacking-health-how-consumers-use-smartphones-and-wearable-tech-to-track-their-health/>.
77. Bol, N., N. Helberger, and J.C. Weert, *Differences in mobile health app use: A source of new digital inequalities?* The Information Society, 2018. **34**(3): p. 183-193.

78. Riley, S., A. Evans, and M. Robson, *Postfeminism and health: Critical psychology and media perspectives*. 2018: Routledge.
79. Pew Research Center. *African Americans and technology use. A demographic portrait*. 2014 5 May 2016]; Available from: <http://www.pewinternet.org/files/2014/01/African-Americans-and-Technology-Use.pdf>. Published 2014.
80. Office for National Statistics. *Internet access – households and individuals: 2017*. 2017; Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017>.
81. Research2Guidance, *Mhealth App Economics 2017: Current Status and Future Trends in Mobile Health*. 2017, Research 2 Guidance Berlin.
82. Campbell, D., *NHS to offer free devices and apps to help people manage illnesses*. The Guardian, 2016.
83. BinDhim, N.F., A. Hawkey, and L. Trevena, *A systematic review of quality assessment methods for smartphone health apps*. *Telemed J E Health*, 2015. **21**(2): p. 97-104.
84. Direito, A., et al., *mHealth technologies to influence physical activity and sedentary behaviors: Behavior change techniques, systematic review and meta-analysis of randomized controlled trials*. *Annals of behavioral medicine*, 2017. **51**(2): p. 226-239.
85. Newton, J. and B. Ferguson. *Making the economic case for prevention*. *Public health matters* 2017; Available from: <https://publichealthmatters.blog.gov.uk/2017/09/06/making-the-economic-case-for-prevention/>.
86. Steel, N., et al., *Changes in health in the countries of the UK and 150 English Local Authority areas 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016*. *The Lancet*, 2018. **392**(10158): p. 1647-1661.
87. Milton, K. and A. Bauman, *A critical analysis of the cycles of physical activity policy in England*. *International Journal of Behavioral Nutrition and Physical Activity*, 2015. **12**(1): p. 8.
88. Stanner, S., *At Least Five a Week—a summary of the report from the Chief Medical Officer on physical activity*. *Nutrition Bulletin*, 2004. **29**(4): p. 350-352.
89. Active, S.A.S. *Start Active, Stay Active: A Report on Physical Activity for Health from the Four Home countries' Chief Medical Officers*. 2011; Available from: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/216370/dh\\_128210.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216370/dh_128210.pdf).
90. ukactive, *ukactive's Blueprint for an Active Britain*. 2018.
91. Organization, W.H., *Global action plan on physical activity 2018-2030: more active people for a healthier world*. 2019: World Health Organization.
92. Government, H., *Sporting future: a new strategy for an active nation*. 2015, Cabinet Office London.
93. Department of Transport, *Cycling and Walking Investment Strategy*. 2017.
94. Stead, M., et al., *Public Health Research, in Mass media to communicate public health messages in six health topic areas: a systematic review and other reviews of the evidence*. 2019, NIHR Journals Library

Copyright (c) Queen's Printer and Controller of HMSO 2019. This work was produced by Stead et al. under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be



addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.: Southampton (UK).

95. Lang, T. and G. Rayner, *Overcoming policy cacophony on obesity: an ecological public health framework for policymakers*. *Obesity reviews*, 2007. **8**: p. 165-181.
96. Walls, H.L., et al., *Public health campaigns and obesity-a critique*. *BMC public health*, 2011. **11**(1): p. 136.
97. Hancock, M., *'My priorities for the health and social care system*. Speech 20th July, 2018.
98. England, P.H., *Improving people's health: applying behavioural and social sciences*. 2018.
99. Rosenstock, I.M., *Historical origins of the health belief model*. *Health education monographs*, 1974. **2**(4): p. 328-335.
100. Bandura, A. and R.H. Walters, *Social learning theory*. Vol. 1. 1977: Prentice-hall Englewood Cliffs, NJ.
101. Velicer, W., et al., *Detailed overview of the transtheoretical model*. *Homeostasis*, 1998. **38**: p. 216-33.
102. Ajzen, I., *The theory of planned behavior*. *Organizational behavior and human decision processes*, 1991. **50**(2): p. 179-211.
103. Webb, T.L. and P. Sheeran, *Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence*. *Psychological bulletin*, 2006. **132**(2): p. 249.
104. Davis, R., et al., *Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review*. *Health psychology review*, 2015. **9**(3): p. 323-344.
105. Houlihan, S., *Dual-process models of health-related behaviour and cognition: a review of theory*. *Public health*, 2018. **156**: p. 52-59.
106. Marteau, T., *Changing minds about changing behaviour*. 2018.
107. Craig, P., et al., *Developing and evaluating complex interventions: the new Medical Research Council guidance*. *Bmj*, 2008. **337**: p. a1655.
108. Lewis, B.A., et al., *Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination*. *Journal of behavioral medicine*, 2017. **40**(1): p. 112-126.
109. Ekkekakis, P., M.E. Hartman, and M.A. Ladwig, *Conceptual foundations of exercise psychology: Facilitators, inhibitors, and a road map toward establishing societal relevance*, in *APA handbook of sport and exercise psychology, volume 2: Exercise psychology, Vol. 2*. 2019, American Psychological Association: Washington, DC, US. p. 27-56.
110. West, R., et al., *Development of a formal system for representing behaviour-change theories*. *Nature human behaviour*, 2019. **3**(5): p. 526.
111. Sallis, J.F., N. Owen, and E. Fisher, *Ecological models of health behavior*. *Health behavior: Theory, research, and practice*, 2015. **5**: p. 43-64.
112. Atkins, L., S. Michie, and C. Wood, *Applying evidence-based theory to behavioural interventions*. Wessex Public Health Network CPD Event: Behaviour Change Management, 2014. **25**.
113. Michie, S., L. Atkins, and R. West, *The Behaviour Change Wheel: A Guide to Designing Interventions*. 2014, GB: Silverback Publishing
114. Michie, S., M.M. van Stralen, and R. West, *The behaviour change wheel: A new method for characterising and designing behaviour change interventions*. *Implementation Science*, 2011. **6**(1): p. 42.
115. Francis, J.J., D. O'Connor, and J. Curran, *Theories of behaviour change synthesised into a set of theoretical groupings: introducing a thematic series on the theoretical domains framework*. *Implementation Science*, 2012. **7**(1): p. 35.



116. Michie, S., et al., *The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions*. *Annals of behavioral medicine*, 2013. **46**(1): p. 81-95.
117. Bardus, M., et al., *Mobile Phone and Web 2.0 Technologies for Weight Management: A Systematic Scoping Review*. *J Med Internet Res*, 2015. **17**(11): p. e259.
118. Campbell, S.M., M.O. Roland, and S.A. Buetow, *Defining quality of care*. *Social science & medicine*, 2000. **51**(11): p. 1611-1625.
119. Maxwell, R.J., *Quality assessment in health*. *Br Med J (Clin Res Ed)*, 1984. **288**(6428): p. 1470-1472.
120. Donabedian, A., *The quality of care: How can it be assessed?* *Jama*, 1988. **260**(12): p. 1743-1748.
121. Department of Health. *High quality care for all: NHS next stage review final report*. London: Department of Health 2008; Available from: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/228836/7432.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf).
122. Bull, F., et al., *Physical activity guidelines in the UK: review and recommendations*. School of Sport, Exercise and Health Sciences, Loughborough University, 2010.
123. Health, U.D.o. and H. Services, *Physical activity guidelines advisory committee*. Washington DC: US Department of Health and Human Services, 2008.
124. Powell, A.C., A.B. Landman, and D.W. Bates, *In search of a few good apps*. *JAMA*, 2014. **311**(18): p. 1851-2.
125. McMillan, B., et al., *Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance*. *Patient Educ Couns*, 2016. **99**(3): p. 429-35.
126. Wicks, P. and E. Chiauzzi, *'Trust but verify'--five approaches to ensure safe medical apps*. *BMC Med*, 2015. **13**: p. 205.
127. Information Commissioner's Office. *Privacy in mobile apps. Guidance for app developers*. 2013 26/2/17; Available from: <https://ico.org.uk/media/for-organisations/documents/1596/privacy-in-mobile-apps-dp-guidance.pdf>.
128. DIRECTIVE, H.A.T., *DIRECTIVE 2009/139/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. 2009.
129. Flores Mateo, G., et al., *Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis*. *J Med Internet Res*, 2015. **17**(11): p. e253.
130. Armstrong, S., *Which app should I use?* *BMJ*, 2015. **351**: p. h4597.
131. National Institute for Health and Care Excellence. *Behavior Change: Individual Approaches*. 2014 5 May 2016 Available from: <https://www.nice.org.uk/guidance/ph49/chapter/What-is-this-guidance-about>.
132. McMillan, B., et al., *The need for quality assurance of health apps*. *BMJ*, 2015. **351**: p. h5915.
133. Boudreaux, E.D., et al., *Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations*. *Transl Behav Med*, 2014. **4**(4): p. 363-71.
134. Glanz, K. and D.B. Bishop, *The role of behavioral science theory in development and implementation of public health interventions*. *Annual review of public health*, 2010. **31**: p. 399-418.
135. Michie, S. and A. Prestwich, *Are interventions theory-based? Development of a theory coding scheme*. *Health Psychology*, 2010. **29**(1): p. 1.
136. Bardus, M., et al., *A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular*

- commercial apps for weight management*. Int J Behav Nutr Phys Act, 2016. **13**: p. 35.
137. Chen, J., J.E. Cade, and M. Allman-Farinelli, *The Most Popular Smartphone Apps for Weight Loss: A Quality Assessment*. JMIR Mhealth Uhealth, 2015. **3**(4): p. e104.
  138. Crane, D., et al., *Behavior change techniques in popular alcohol reduction apps: content analysis*. J Med Internet Res, 2015. **17**(5): p. e118.
  139. Yang, C.H., J.P. Maher, and D.E. Conroy, *Implementation of behavior change techniques in mobile applications for physical activity*. Am J Prev Med, 2015. **48**(4): p. 452-5.
  140. Michie, S., et al., *Effective techniques in healthy eating and physical activity interventions: a meta-regression*. Health Psychology, 2009. **28**(6): p. 690.
  141. Pearson, E.S., *Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components*. Patient education and counseling, 2012. **87**(1): p. 32-42.
  142. Lustria, M.L.A., et al., *A meta-analysis of web-delivered tailored health behavior change interventions*. Journal of health communication, 2013. **18**(9): p. 1039-1069.
  143. Hankonen, N., et al., *Which behavior change techniques are associated with changes in physical activity, diet and body mass index in people with recently diagnosed diabetes?* Annals of Behavioral Medicine, 2014. **49**(1): p. 7-17.
  144. Abraham, C. and S. Michie, *A taxonomy of behavior change techniques used in interventions*. Health psychology, 2008. **27**(3): p. 379.
  145. Middelweerd, A., et al., *Apps to promote physical activity among adults: a review and content analysis*. Int J Behav Nutr Phys Act, 2014. **11**(1): p. 97.
  146. Harman, M., Y. Jia, and Y. Zhang. *App store mining and analysis: MSR for app stores*. in *Proceedings of the 9th IEEE Working Conference on Mining Software Repositories*. 2012. IEEE Press.
  147. Mendiola, M.F., M. Kalnicki, and S. Lindenauer, *Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings*. JMIR Mhealth Uhealth, 2015. **3**(2): p. e40.
  148. Girardello, A. and F. Michahelles. *AppAware: which mobile applications are hot?* in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 2010. ACM.
  149. Kuehnhausen, M. and V.S. Frost. *Trusting smartphone apps? To install or not to install, that is the question*. in *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2013 IEEE International Multi-Disciplinary Conference on*. 2013. IEEE.
  150. Al-Subaihini, A., et al., *App store mining and analysis*. 2015: p. 1-2.
  151. Eysenbach, G., *What is e-health?* J Med Internet Res, 2001. **3**(2): p. e20.
  152. Yardley, L., et al., *Understanding and Promoting Engagement with Digital Health Interventions*. in press.
  153. Powell, A.C., et al., *Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps*. JMIR Mhealth Uhealth, 2016. **4**(1): p. e15.
  154. Nielsen, J., *Usability engineering*. 1994: Elsevier.
  155. Nielsen, J., *Heuristic evaluation*. Usability inspection methods, 1994. **17**(1): p. 25-62.
  156. Kratzke, C. and C. Cox, *Smartphone technology and apps: rapidly changing health promotion*. Global Journal of Health Education and Promotion, 2012. **15**(1).
  157. Stoyanov, S.R., et al., *Mobile app rating scale: a new tool for assessing the quality of health mobile apps*. JMIR mHealth and uHealth, 2015. **3**(1).

158. *Oxford dictionary of English*. 2nd ed. / edited by Catherine Soanes and Angus Stevenson. ed, ed. C. Soanes and A. Stevenson. 2003, Oxford: Oxford : Oxford University Press.
159. Deterding, S., et al., *From Game Design Elements to Gamefulness: Defining "Gamification"*. 2011.
160. Lister, C., et al., *Just a fad? Gamification in health and fitness apps*. JMIR serious games, 2014. **2**(2): p. e9.
161. Brooke, J., *SUS-A quick and dirty usability scale*. Usability evaluation in industry, 1996. **189**(194): p. 4-7.
162. Brooke, J., *System usability scale (SUS): a quick-and-dirty method of system evaluation user information*. Reading, UK: Digital Equipment Co Ltd, 1986.
163. Singh, K., et al., *Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications*. Issue Brief (Commonw Fund), 2016. **5**: p. 1-11.
164. Kirakowski, J., N. Claridge, and R. Whitehand. *Human centered measures of success in Web design*. in *Conference Proceedings of the 4th Conference on Human Factors and the Web*. 1998.
165. Bangor, A., P.T. Kortum, and J.T. Miller, *An empirical evaluation of the system usability scale*. Intl. Journal of Human-Computer Interaction, 2008. **24**(6): p. 574-594.
166. Lewis, J.R., *Psychometric evaluation of the PSSUQ using data from five years of usability studies*. International Journal of Human-Computer Interaction, 2002. **14**(3-4): p. 463-488.
167. Conroy, D.E., C.H. Yang, and J.P. Maher, *Behavior change techniques in top-ranked mobile apps for physical activity*. Am J Prev Med, 2014. **46**(6): p. 649-52.
168. Michie, S., et al., *A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy*. Psychology & Health, 2011. **26**(11): p. 1479-1498.
169. Cowan, L.T., et al., *Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory*. Health Educ Behav, 2013. **40**(2): p. 133-9.
170. Knight, E., et al., *Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores*. JMIR Mhealth Uhealth, 2015. **3**(2): p. e43.
171. West, J.H., et al., *There's an app for that: content analysis of paid health and fitness apps*. J Med Internet Res, 2012. **14**(3): p. e72.
172. Azar, K.M., et al., *Mobile applications for weight management: theory-based content analysis*. Am J Prev Med, 2013. **45**(5): p. 583-9.
173. Breton, E.R., B.F. Fuemmeler, and L.C. Abrams, *Weight loss-there is an app for that! But does it adhere to evidence-informed practices?* Transl Behav Med, 2011. **1**(4): p. 523-9.
174. Hodgdon, M. *Value of organic first-page results 2015* [cited 2016 12/07/2016]; Available from: <http://www.infront.com/blogs/the-infront-blog/2015/6/17/value-of-first-page-google-results>.
175. NHS Choices. *Physical activity guidelines for adults*. Live Well 2015 [cited 2016 21 May ]; Available from: <http://www.nhs.uk/Livewell/fitness/Pages/physical-activity-guidelines-for-adults.aspx>.
176. Direito, A., et al., *Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques?* BMC Public Health, 2014. **14**(1): p. 1.
177. Stoyanov, S.R., et al., *Mobile app rating scale: a new tool for assessing the quality of health mobile apps*. JMIR Mhealth Uhealth, 2015. **3**(1): p. e27.

178. Mendiola, M.F., M. Kalnicki, and S. Lindenauer, *Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings*. JMIR mHealth and uHealth, 2015. **3**(2).
179. Finstad, K., *The system usability scale and non-native english speakers*. Journal of usability studies, 2006. **1**(4): p. 185-188.
180. Nielsen, J., *Usability 101: Introduction to usability*. 2003.
181. Cohen, J., *Kappa: Coefficient of concordance*. Educ. Psych. Measurement, 1960. **20**(37).
182. Byrt, T., J. Bishop, and J.B. Carlin, *Bias, prevalence and kappa*. Journal of clinical epidemiology, 1993. **46**(5): p. 423-429.
183. Cunningham, M. *More than just the kappa coefficient: a program to fully characterize inter-rater reliability between two raters*. in *SAS global forum*. 2009.
184. Ubhi, H.K., et al., *Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features*. Translational behavioral medicine, 2016. **6**(3): p. 410-417.
185. Klika, B. and C. Jordan, *High-intensity circuit training using body weight: Maximum results with minimal investment*. ACSM's Health & Fitness Journal, 2013. **17**(3): p. 8-13.
186. Deterding, S., et al. *From game design elements to gamefulness: defining gamification*. in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. 2011. ACM.
187. Direito, A., et al., *Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial*. Journal of medical Internet research, 2015. **17**(8).
188. Conn, V.S., A.R. Hafdahl, and D.R. Mehr, *Interventions to increase physical activity among healthy adults: meta-analysis of outcomes*. American journal of public health, 2011. **101**(4): p. 751-758.
189. Marteau, T.M., G.J. Hollands, and P.C. Fletcher, *Changing human behavior to prevent disease: the importance of targeting automatic processes*. science, 2012. **337**(6101): p. 1492-1495.
190. van Genugten, L., et al., *Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis*. Journal of medical Internet research, 2016. **18**(6): p. e155.
191. Webb, T., et al., *Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy*. Journal of medical Internet research, 2010. **12**(1): p. e4.
192. Subhi, Y., et al., *Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review*. JMIR mHealth and uHealth, 2015. **3**(3): p. e79.
193. Pereira-Azevedo, N., et al., *Expert involvement predicts mHealth app downloads: multivariate regression analysis of urology apps*. JMIR mHealth and uHealth, 2016. **4**(3).
194. Martínez-Pérez, B., I. De La Torre-Díez, and M. López-Coronado, *Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis*. Journal of medical Internet research, 2013. **15**(6).
195. Modave, F., et al., *Low quality of free coaching apps with respect to the American College of Sports Medicine guidelines: a review of current mobile apps*. JMIR mHealth and uHealth, 2015. **3**(3).
196. adjust, *The Zombie Uprising. A look at the undead App Store in 2016*. 2016.
197. Yardley, L., et al., *Understanding and promoting effective engagement with digital behavior change interventions*. American journal of preventive medicine, 2016. **51**(5): p. 833-842.

198. Waltz, A., *Deconstructing the App Store Rankings Formula with a Little Mad Science*. 2015, Moz.
199. National Information Board. *Personalised Health Care 2020: Work stream 1.2 Roadmap. Enable me to make the right health and care choices. Providing citizens with access to an assessed set of NHS and social care 'apps'*. 2015; Available from: <https://www.gov.uk/government/publications/national-information-boards-workstreams>.
200. Donkin, L., et al., *A systematic review of the impact of adherence on the effectiveness of e-therapies*. Journal of medical Internet research, 2011. **13**(3).
201. Pustozarov, E., U. von Jan, and U.V. Albrecht, *Evaluation of mHealth Applications Quality Based on User Ratings*. Stud Health Technol Inform, 2016. **226**: p. 237-40.
202. Business Insider. *BlackBerry's share of the global smartphone market is now officially 0%*. 2017 [cited 2017; Available from: <http://uk.businessinsider.com/blackberry-smartphone-marketshare-zero-percent-gartner-q4-2016-2017-2>.
203. Arnhold, M., M. Quade, and W. Kirch, *Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older*. J Med Internet Res, 2014. **16**(4): p. e104.
204. Deci, E.L. and R.M. Ryan, *Self-determination theory*. Handbook of theories of social psychology, 2011. **1**(2011): p. 416-433.
205. Riley, W.T., et al., *Health behavior models in the age of mobile interventions: are our theories up to the task?* Translational behavioral medicine, 2011. **1**(1): p. 53-71.
206. Abrams, L.C., et al., *iPhone apps for smoking cessation: a content analysis*. American journal of preventive medicine, 2011. **40**(3): p. 279-285.
207. Greaves, C.J., et al., *Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions*. BMC public health, 2011. **11**(1): p. 119.
208. Carver, C.S. and M.F. Scheier, *Control theory: A useful conceptual framework for personality–social, clinical, and health psychology*. Psychological bulletin, 1982. **92**(1): p. 111.
209. Kanfer, F.H., *Implications of a self-regulation model of therapy for treatment of addictive behaviors*, in *Treating addictive behaviors*. 1986, Springer. p. 29-47.
210. Greenland, S., J. Pearl, and J.M. Robins, *Causal diagrams for epidemiologic research*. Epidemiology, 1999: p. 37-48.
211. Shrier, I. and R.W. Platt, *Reducing bias through directed acyclic graphs*. BMC medical research methodology, 2008. **8**(1): p. 70.
212. Dupont, W.D. and W.D. Plummer Jr, *Power and sample size calculations for studies involving linear regression*. Controlled clinical trials, 1998. **19**(6): p. 589-601.
213. Ara, R., et al., *Predicting preference-based utility values using partial proportional odds models*. BMC Res Notes, 2014. **7**: p. 438.
214. Field, A., *Discovering statistics using IBM SPSS statistics*. 2013: sage.
215. George, D., *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e*. 2011: Pearson Education India.
216. Bangor, A., P. Kortum, and J. Miller, *Determining what individual SUS scores mean: Adding an adjective rating scale*. Journal of usability studies, 2009. **4**(3): p. 114-123.
217. Wood, C.E., et al., *Applying the behaviour change technique (BCT) taxonomy v1: a study of coder training*. Translational behavioral medicine, 2014. **5**(2): p. 134-148.



218. Tiberius, V. and A. Hall, *Normative theory and psychological research: Hedonism, eudaimonism, and why it matters*. The Journal of Positive Psychology, 2010. **5**(3): p. 212-225.
219. Kashdan, T.B., R. Biswas-Diener, and L.A. King, *Reconsidering happiness: The costs of distinguishing between hedonics and eudaimonia*. The Journal of Positive Psychology, 2008. **3**(4): p. 219-233.
220. Waterman, A.S., *Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment*. Journal of personality and social psychology, 1993. **64**(4): p. 678.
221. Ickin, S., K. Petersen, and J. Gonzalez-Huerta. *Why Do Users Install and Delete Apps? A Survey Study*. in *International Conference of Software Business*. 2017. Springer.
222. Adams, M.A., et al., *Adaptive goal setting and financial incentives: a 2x2 factorial randomized controlled trial to increase adults' physical activity*. BMC public health, 2017. **17**(1): p. 286.
223. Perski, O., et al., *Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis*. Translational behavioral medicine, 2016. **7**(2): p. 254-267.
224. Weinberg, R.S., *Goal setting in sport and exercise: Research to practice*. 2002.
225. Locke, E.A. and G.P. Latham, *Building a practically useful theory of goal setting and task motivation: A 35-year odyssey*. American psychologist, 2002. **57**(9): p. 705.
226. Götz, F.M., S. Stieger, and U.-D. Reips, *Users of the main smartphone operating systems (iOS, Android) differ only little in personality*. PloS one, 2017. **12**(5): p. e0176921.
227. Monetate. *Ecommerce Quarterly 2016: Mobile ate the world and is still hungry*. 2017 [cited 2018 21/09/18]; Available from: [http://info.monetate.com/rs/092-TQN-434/images/Monetate-EQ3-2016.pdf?mkt\\_tok=eyJpIjoiTURFMFptTTNNR1prWWpRMylsInQiOiJsaDE4bXF1OTA4R0o4T0xiNGZjXC9JRkcxHM4WkRTZFZkt2tnWWorMGtIc24zSUFsanN2Z2tOVDJHTDZFc3F1eTZnZEZJaWJsQUxzQWRjN3F6ZXVnc2wrcVREMDFJc3RMQU5yc0NGdXU4OG41Zk9Gdk1OaGZRUXNEcisreWFEczcifQ%3D%3D](http://info.monetate.com/rs/092-TQN-434/images/Monetate-EQ3-2016.pdf?mkt_tok=eyJpIjoiTURFMFptTTNNR1prWWpRMylsInQiOiJsaDE4bXF1OTA4R0o4T0xiNGZjXC9JRkcxHM4WkRTZFZkt2tnWWorMGtIc24zSUFsanN2Z2tOVDJHTDZFc3F1eTZnZEZJaWJsQUxzQWRjN3F6ZXVnc2wrcVREMDFJc3RMQU5yc0NGdXU4OG41Zk9Gdk1OaGZRUXNEcisreWFEczcifQ%3D%3D).
228. Dick, J. *How to Tell an Android User from an iOS User*. 2013 [cited 2018; Available from: <https://civicscience.com/how-to-tell-an-android-user-from-an-ios-user/>].
229. Kooti, F., et al. *iPhone's Digital Marketplace: Characterizing the Big Spenders*. in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 2017. ACM.
230. Katariya, J. *Apple Vs Android—A comparative study 2017*. 2017 [cited 2017 2018]; Available from: <https://android.jlelse.eu/apple-vs-android-a-comparative-study-2017-c5799a0a1683>.
231. Shaw, H., et al. *Individual differences between iPhone and Android smartphone users*. 2016. British Psychological Society.
232. Hixon, T. *What Kind Of Person Prefers An iPhone?* Forbes 2014 [cited 2018 2018]; Available from: <https://www.forbes.com/sites/toddhixon/2014/04/10/what-kind-of-person-prefers-an-iphone/#5a96dd55d1b0>.
233. Szabolcs, M. *Why do iOS Apps tend to be higher quality than Android Apps?* Quora 2015 [cited 2018; Available from: Why do iOS Apps tend to be higher quality than Android Apps?
234. Guzman, E. and W. Maalej. *How do users like this feature? a fine grained sentiment analysis of app reviews*. in *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*. 2014. IEEE.

235. Pagano, D. and W. Maalej. *User feedback in the appstore: An empirical study*. in *Requirements Engineering Conference (RE), 2013 21st IEEE International*. 2013. IEEE.
236. Buechi, R., et al., *Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis*. *BMJ open*, 2017. **7**(12): p. e018280.
237. Jake-Schoffman, D.E., et al., *Methods for evaluating the content, usability, and efficacy of commercial mobile health apps*. *JMIR mHealth and uHealth*, 2017. **5**(12).
238. Xu, W. and Y. Liu, *mHealthApps: a repository and database of mobile health apps*. *JMIR mHealth and uHealth*, 2015. **3**(1): p. e28.
239. West, R., et al., *Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes*. *Nicotine & Tobacco Research*, 2010. **12**(7): p. 742-747.
240. research2guidance. *mHealth app developer economics. The state of the art of mHealth application publishing*. 2014 5 May 2016; Available from: <http://research2guidance.com/product/mhealth-app-developer-economics-2014/>.
241. Public Health England. *Health app developers: the assessment process. How developers can get their health app assessed and endorsed*. 2017 [cited 2018 2018]; Available from: <https://www.gov.uk/guidance/health-app-developers-the-assessment-process>.
242. Boutron, I., et al., *Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments*. *Annals of internal medicine*, 2008. **148**(4): p. W-60-W-66.
243. Akobeng, A., *Understanding randomised controlled trials*. *Archives of disease in childhood*, 2005. **90**(8): p. 840-844.
244. Feeley, N., et al., *The importance of piloting an RCT intervention*. *CJNR (Canadian Journal of Nursing Research)*, 2009. **41**(2): p. 84-99.
245. Higgins, J.P., et al., *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials*. *Bmj*, 2011. **343**: p. d5928.
246. Richard, M., et al., *Intrinsic motivation and exercise adherence*. *Int J Sport Psychol*, 1997. **28**(4): p. 335-354.
247. Ntoumanis, N., *A self-determination approach to the understanding of motivation in physical education*. *British journal of educational psychology*, 2001. **71**(2): p. 225-242.
248. Baquet, C.R., et al., *Recruitment and participation in clinical trials: socio-demographic, rural/urban, and health care access predictors*. *Cancer detection and prevention*, 2006. **30**(1): p. 24-33.
249. Brueton, V., et al., *Best practice guidance for the use of strategies to improve retention in randomized trials developed from two consensus workshops*. *Journal of clinical epidemiology*, 2017. **88**: p. 122-132.
250. Brueton, V., et al., *Use of strategies to improve retention in primary care randomised trials: a qualitative study with in-depth interviews*. *BMJ open*, 2014. **4**(1): p. e003835.
251. van Hees, V., *The challenge of assessing physical activity in populations*. *The Lancet*, 2012. **380**(9853): p. 1555.
252. Sylvia, L.G., et al., *Practical guide to measuring physical activity*. *Journal of the Academy of Nutrition and Dietetics*, 2014. **114**(2): p. 199-208.
253. Garatachea, N., G. Torres Luque, and J. Gonzalez Gallego, *Physical activity and energy expenditure measurements using accelerometers in older adults*. *Nutrición hospitalaria*, 2010. **25**(2).

254. Miyachi, M., et al., *Installation of a stationary high desk in the workplace: effect of a 6-week intervention on physical activity*. BMC public health, 2015. **15**(1): p. 368.
255. Pears, S., et al., *A randomised controlled trial of three very brief interventions for physical activity in primary care*. BMC Public Health, 2016. **16**(1): p. 1033.
256. Wang, J.B., et al., *Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial*. Telemedicine and e-Health, 2015. **21**(10): p. 782-792.
257. Rhodes, R.E., et al., *Physical activity: Health impact, prevalence, correlates and interventions*. Psychology & Health, 2017. **32**(8): p. 942-975.
258. Rudolf, K., et al., *Impact of different recruitment strategies on accelerometry adherence and resulting physical activity data: A secondary analysis*. Preventive medicine reports, 2018. **10**: p. 76-81.
259. Wijndaele, K., et al., *Utilization and harmonization of adult accelerometry data: review and expert consensus*. Medicine and science in sports and exercise, 2015. **47**(10): p. 2129.
260. Trost, S.G., K.L. Mciver, and R.R. Pate, *Conducting accelerometer-based activity assessments in field-based research*. Medicine and science in sports and exercise, 2005. **37**(11 Suppl): p. S531-43.
261. Hallal, P. and C. Victora, *Reliability and validity of the international physical activity questionnaire (IPAQ)*. Medicine & Science in Sports & Exercise, 2004. **36**(3).
262. Misra, P., et al., *A community based study to test the reliability and validity of physical activity measurement techniques*. International journal of preventive medicine, 2014. **5**(8): p. 952.
263. Bauman, A.E., et al., *Correlates of physical activity: why are some people physically active and others not?* The lancet, 2012. **380**(9838): p. 258-271.
264. Ouellette, J.A. and W. Wood, *Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior*. Psychological bulletin, 1998. **124**(1): p. 54.
265. Schwarzer, R. and R. Fuchs, *Self-efficacy and health behaviours*. Predicting health behavior: Research and practice with social cognition models, 1996. **163**: p. 196.
266. Lippke, S., *Outcome Expectation*, in *Encyclopedia of Personality and Individual Differences*, V. Zeigler-Hill and T.K. Shackelford, Editors. 2017, Springer International Publishing: Cham. p. 1-2.
267. Bandura, A., *Self-efficacy: toward a unifying theory of behavioral change*. Psychological review, 1977. **84**(2): p. 191.
268. Prochaska, J.O. and C.C. DiClemente, *Transtheoretical therapy: Toward a more integrative model of change*. Psychotherapy: theory, research & practice, 1982. **19**(3): p. 276.
269. Williams, D.M., E.S. Anderson, and R.A. Winett, *A review of the outcome expectancy construct in physical activity research*. Annals of behavioral medicine, 2005. **29**(1): p. 70-79.
270. Higgins, J., ed. *Revised Cochrane risk of bias tool for randomized trials (RoB 2.0). Additional considerations for cross-over trials*. 2016.
271. Lally, P., et al., *How are habits formed: Modelling habit formation in the real world*. European journal of social psychology, 2010. **40**(6): p. 998-1009.
272. Perski, O., et al., *A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale"*. Translational Behavioral Medicine, 2019.



273. Scarborough, P., et al., *The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs*. Journal of public health, 2011. **33**(4): p. 527-535.
274. Mackert, M., et al., *Health literacy and health information technology adoption: the potential for a new digital divide*. Journal of medical Internet research, 2016. **18**(10).
275. Tucker, P. and J. Gilliland, *The effect of season and weather on physical activity: a systematic review*. Public health, 2007. **121**(12): p. 909-922.
276. Chan, C. and D. Ryan, *Assessing the effects of weather conditions on physical activity participation using objective measures*. International journal of environmental research and public health, 2009. **6**(10): p. 2639-2654.
277. Eldridge, S.M., et al., *CONSORT 2010 statement: extension to randomised pilot and feasibility trials*. Pilot and feasibility studies, 2016. **2**(1): p. 64.
278. Eysenbach, G. and C.-E. Group, *CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions*. Journal of medical Internet research, 2011. **13**(4).
279. Khadjesari, Z., et al., *Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials*. Journal of Medical Internet Research, 2011. **13**(1).
280. Linke, S., et al., *Internet-based interactive health intervention for the promotion of sensible drinking: patterns of use and potential impact on members of the general public*. Journal of Medical Internet Research, 2007. **9**(2).
281. Hoffmann, T.C., et al., *Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide*. Bmj, 2014. **348**: p. g1687.
282. Harris, T.J., et al., *What factors are associated with physical activity in older people, assessed objectively by accelerometry?* British journal of sports medicine, 2009. **43**(6): p. 442-450.
283. Davis, M.G., et al., *Objectively measured physical activity in a diverse sample of older urban UK adults*. Medicine & Science in Sports & Exercise, 2011. **43**(4): p. 647-654.
284. Hart, T.L., et al., *How many days of monitoring predict physical activity and sedentary behaviour in older adults?* International Journal of Behavioral Nutrition and Physical Activity, 2011. **8**(1): p. 62.
285. Jefferis, B.J., et al., *Adherence to physical activity guidelines in older adults, using objectively measured physical activity in a population-based study*. BMC Public Health, 2014. **14**(1): p. 382.
286. Choi, L., et al., *Validation of accelerometer wear and nonwear time classification algorithm*. Medicine and science in sports and exercise, 2011. **43**(2): p. 357.
287. Aibar, A., et al., *Do epoch lengths affect adolescents' compliance with physical activity guidelines?* The Journal of sports medicine and physical fitness, 2014. **54**(3): p. 326-334.
288. Freedson, P.S., E. Melanson, and J. Sirard, *Calibration of the Computer Science and Applications, Inc. accelerometer*. Medicine and science in sports and exercise, 1998. **30**(5): p. 777-781.
289. IPAQ. *Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short and Long Forms, revised on November 2005*. 2005 15/01/18]; Available from: <http://www.ipaq.ki.se/scoring.pdf>.
290. Steinhardt, M.A. and R.K. Dishman, *Reliability and validity of expected outcomes and barriers for habitual physical activity*. Journal of occupational medicine.: official publication of the Industrial Medical Association, 1989. **31**(6): p. 536-546.

291. Bandura, A., *Guide for constructing self-efficacy scales*. Self-efficacy beliefs of adolescents, 2006. **5**(307-337).
292. Chan, J.S.Y., et al., *Special Issue - Therapeutic Benefits of Physical Activity for Mood: A Systematic Review on the Effects of Exercise Intensity, Duration, and Modality*. J Psychol, 2019. **153**(1): p. 102-125.
293. Peluso, M.A. and L.H. Guerra de Andrade, *Physical activity and mental health: the association between exercise and mood*. Clinics (Sao Paulo), 2005. **60**(1): p. 61-70.
294. Yeung, R.R., *The acute effects of exercise on mood state*. Journal of psychosomatic research, 1996. **40**(2): p. 123-141.
295. Paluska, S.A. and T.L. Schwenk, *Physical activity and mental health*. Sports medicine, 2000. **29**(3): p. 167-180.
296. Koval, P., et al., *Affective instability in daily life is predicted by resting heart rate variability*. PLoS One, 2013. **8**(11): p. e81536.
297. Rodrigues, A.M., et al., *The question–behavior effect: Genuine effect or spurious phenomenon? A systematic review of randomized controlled trials with meta-analyses*. Health Psychology, 2015. **34**(1): p. 61.
298. McCambridge, J. and K. Kypri, *Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials*. PloS one, 2011. **6**(10): p. e23748.
299. Gail, M.H., et al., *Encyclopedia of epidemiologic methods*. 2000: John Wiley & Sons.
300. Hooper, R. *Justifying sample size for a feasibility study*.
301. Brown, T.R. and V.I. Simnad, *A randomized crossover trial of dalfampridine extended release for effect on ambulatory activity in people with multiple sclerosis*. International journal of MS care, 2016. **18**(4): p. 170-176.
302. Schulz, K.F., D.G. Altman, and D. Moher, *CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials*. BMC medicine, 2010. **8**(1): p. 18.
303. Gomes, M., et al., *Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance*. Health economics, 2016. **25**(5): p. 515-528.
304. *PaedePPOC Clinical Reference Manual*, A.h.s.r. institute, Editor. 2012.
305. Kyu, H.H., et al., *Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013*. bmj, 2016. **354**: p. i3857.
306. Rzewnicki, R., Y.V. Auweele, and I. De Bourdeaudhuij, *Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample*. Public health nutrition, 2003. **6**(3): p. 299-305.
307. De Cocker, K., et al., *Self-reported physical activity in European adolescents: results from the HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) study*. Public health nutrition, 2011. **14**(2): p. 246-254.
308. Guse, K., et al., *Interventions using new digital media to improve adolescent sexual health: a systematic review*. Journal of adolescent health, 2012. **51**(6): p. 535-543.
309. Marin, T.S., et al., *Examining adherence to activity monitoring devices to improve physical activity in adults with cardiovascular disease: A systematic review*. European journal of preventive cardiology, 2019. **26**(4): p. 382-397.
310. Jefferis, B.J., et al., *Objectively measured physical activity, sedentary behaviour and all-cause mortality in older men: does volume of activity matter more than pattern of accumulation?* Br J Sports Med, 2018: p. bjsports-2017-098733.

311. Pedišić, Ž. and A. Bauman, *Accelerometer-based measures in physical activity surveillance: current practices and issues*. Br J Sports Med, 2015. **49**(4): p. 219-223.
312. Kwasnicka, D., et al., *Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories*. Health psychology review, 2016. **10**(3): p. 277-296.
313. Lisspers, J., et al., *Long-term effects of lifestyle behavior change in coronary artery disease: effects on recurrent coronary events after percutaneous coronary intervention*. Health Psychology, 2005. **24**(1): p. 41.
314. Dunton, G.F., et al., *Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity*. Frontiers in psychology, 2012. **3**: p. 260.
315. Prince, S.A., et al., *A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review*. International journal of behavioral nutrition and physical activity, 2008. **5**(1): p. 56.
316. Adamo, K.B., et al., *A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review*. International Journal of Pediatric Obesity, 2009. **4**(1): p. 2-27.
317. Pfister, T., et al., *Comparison of two accelerometers for measuring physical activity and sedentary behaviour*. BMJ open sport & exercise medicine, 2017. **3**(1): p. e000227.
318. Lee, P.H., et al., *Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review*. International Journal of Behavioral Nutrition and Physical Activity, 2011. **8**(1): p. 115.
319. van Poppel, M.N., et al., *Physical activity questionnaires for adults*. Sports medicine, 2010. **40**(7): p. 565-600.
320. Fagaras, S.-P., L.-E. Radu, and G. Vanvu, *The level of physical activity of university students*. Procedia-Social and Behavioral Sciences, 2015. **197**: p. 1454-1457.
321. Mezulis, A.H., et al., *Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias*. Psychological bulletin, 2004. **130**(5): p. 711.
322. Santos, I., et al., *Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis*. Obesity Reviews, 2017. **18**(1): p. 32-50.
323. Harper, H. and M. Hallsworth, *Counting Calories: how under-reporting can explain the apparent fall in calorie intake*. 2016.
324. Rzewnicki, R., Y. Vanden Auweele, and I. De Bourdeaudhuij, *Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample*. Public Health Nutr, 2003. **6**(3): p. 299-305.
325. Bermúdez, V.J., et al., *International physical activity questionnaire overestimation is ameliorated by individual analysis of the scores*. American journal of therapeutics, 2013. **20**(4): p. 448-458.
326. Doherty, K. and G. Doherty, *The construal of experience in HCI: Understanding self-reports*. International Journal of Human-Computer Studies, 2018. **110**: p. 63-74.
327. Reis, H.T., *Why researchers should think "real-world": A conceptual rationale*. 2012.
328. Donkin, L., et al., *A systematic review of the impact of adherence on the effectiveness of e-therapies*. Journal of medical Internet research, 2011. **13**(3): p. e52.
329. Schwarzer, R., *Social-cognitive factors in changing health-related behaviors*. Current directions in psychological science, 2001. **10**(2): p. 47-51.

330. Godin, G., et al., *Asking questions changes behavior: mere measurement effects on frequency of blood donation*. *Health Psychology*, 2008. **27**(2): p. 179.
331. Sandberg, T. and M. Conner, *A mere measurement effect for anticipated regret: Impacts on cervical screening attendance*. *British Journal of Social Psychology*, 2009. **48**(2): p. 221-236.
332. French, D.P. and S. Sutton, *Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it?* *British journal of health psychology*, 2010. **15**(3): p. 453-468.
333. Maher, J.P., et al., *Do fluctuations in positive affective and physical feeling states predict physical activity and sedentary time?* *Psychology of sport and exercise*, 2019. **41**: p. 153-161.
334. Dunton, G.F., et al., *Momentary assessment of affect, physical feeling states, and physical activity in children*. *Health Psychol*, 2014. **33**(3): p. 255-63.
335. Dunton, G.F., et al., *Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones*. *Obesity (Silver Spring)*, 2011. **19**(6): p. 1205-12.
336. Maher, J.P., et al., *A daily analysis of physical activity and satisfaction with life in emerging adults*. *Health Psychol*, 2013. **32**(6): p. 647-56.
337. Young, L., M. Hertzog, and S. Barnason, *Feasibility of Using Accelerometer Measurements to Assess Habitual Physical Activity in Rural Heart Failure Patients*. *Geriatrics*, 2017. **2**(3): p. 23.
338. Kowalski, K., et al., *Direct and indirect measurement of physical activity in older adults: a systematic review of the literature*. *International Journal of Behavioral Nutrition and Physical Activity*, 2012. **9**(1): p. 148.
339. Gal, R., et al., *The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis*. *Sports medicine-open*, 2018. **4**(1): p. 42.
340. *Census 2011*. Census Information Scheme 2011 02/06/19]; Available from: <https://data.london.gov.uk/census/themes/diversity/>.
341. Office for National Statistics. *2011 Census: Key Statistics and Quick Statistics for Local Authorities in the United Kingdom*. 2013 02/06/2019]; Available from: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-12-04#background-notes>.
342. Eurostat. *Tertiary educational attainment, age group 25-64 by sex and NUTS 2 regions*. Available from: <https://ec.europa.eu/eurostat/tgm/graph.do?tab=graph&plugin=1&language=en&pcode=tgs00109&toolbox=type>.
343. Short, C.E., et al., *Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies*. *Journal of medical Internet research*, 2018. **20**(11): p. e292.
344. Ekkekakis, P. and Z. Zenko, *Escape from cognitivism: Exercise as hedonic experience*, in *Sport and exercise psychology research*. 2016, Elsevier. p. 389-414.
345. Moore, G.F., et al., *Process evaluation of complex interventions: Medical Research Council guidance*. *bmj*, 2015. **350**: p. h1258.
346. Nastasi, B.K., et al., *The Participatory Intervention Model: A framework for conceptualizing and promoting intervention acceptability*. *School Psychology Quarterly*, 2000. **15**(2): p. 207.
347. Sekhon, M., M. Cartwright, and J.J. Francis, *Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework*. *BMC health services research*, 2017. **17**(1): p. 88.

348. Kwasnicka, D., et al., *Data-prompted interviews: Using individual ecological data to stimulate narratives and explore meanings*. *Health Psychology*, 2015. **34**(12): p. 1191.
349. Rodriguez, S.K. and M.R. Kerrigan, *Using graphic elicitation to explore community college transfer student identity, development, and engagement*. *The Qualitative Report*, 2016. **21**(6): p. 1052-1070.
350. Arden, M.A., et al., *Adherence to medication in adults with Cystic Fibrosis: An investigation using objective adherence data and the Theoretical Domains Framework*. *British journal of health psychology*, 2019. **24**(2): p. 357-380.
351. Kwasnicka, D., et al., *'It's not a diet, it's a lifestyle': a longitudinal, data-prompted interview study of weight loss maintenance*. *Psychology & health*, 2019: p. 1-20.
352. Hardcastle, S.J., K. McNamara, and L. Tritton, *Using visual methods to understand physical activity maintenance following cardiac rehabilitation*. *PLoS one*, 2015. **10**(9): p. e0138218.
353. Naughton, F., et al., *A context-sensing mobile phone app (Q sense) for smoking cessation: a mixed-methods study*. *JMIR mHealth and uHealth*, 2016. **4**(3): p. e106.
354. Richman, W.L., et al., *A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews*. *Journal of applied psychology*, 1999. **84**(5): p. 754.
355. Patey, A.M., et al., *Anesthesiologists' and surgeons' perceptions about routine pre-operative testing in low-risk patients: application of the Theoretical Domains Framework (TDF) to identify factors that influence physicians' decisions to order pre-operative tests*. *Implementation Science*, 2012. **7**(1): p. 52.
356. Huijg, J.M., et al., *Measuring determinants of implementation behavior: psychometric properties of a questionnaire based on the theoretical domains framework*. *Implementation Science*, 2014. **9**(1): p. 33.
357. Power, B.T., et al., *Understanding perceived determinants of nurses' eating and physical activity behaviour: a theory-informed qualitative interview study*. *BMC obesity*, 2017. **4**(1): p. 18.
358. Seward, K., et al., *Measuring implementation behaviour of menu guidelines in the childcare setting: confirmatory factor analysis of a theoretical domains framework questionnaire (TDFQ)*. *International Journal of Behavioral Nutrition and Physical Activity*, 2017. **14**(1): p. 45.
359. Stevely, A.K., et al., *Exposure to revised drinking guidelines and 'COM-B' determinants of behaviour change: descriptive analysis of a monthly cross-sectional survey in England*. *BMC public health*, 2018. **18**(1): p. 251.
360. McParlin, C., et al., *What helps or hinders midwives to implement physical activity guidelines for obese pregnant women? A questionnaire survey using the Theoretical Domains Framework*. *Midwifery*, 2017. **49**: p. 110-116.
361. Banister, P., *Qualitative methods in psychology: A research guide*. 2011: McGraw-Hill Education (UK).
362. Garnett, C., 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*. *JMIR mHealth and uHealth*, 2018. **6**(12): p. e11175.
363. Linke, S., et al., *Internet-based interactive health intervention for the promotion of sensible drinking: patterns of use and potential impact on members of the general public*. *Journal of Medical Internet Research*, 2007. **9**(2): p. e10.
364. Braun, V. and V. Clarke, *Using thematic analysis in psychology*. *Qualitative research in psychology*, 2006. **3**(2): p. 77-101.
365. Churches, O. and M.E. Nicholls, *Near Misses and the Effect of Attentional Asymmetries on Sporting Performance*, in *Laterality in Sports*. 2016, Elsevier. p. 205-224.

366. Macintyre, S., *The social patterning of exercise behaviours: the role of personal and local resources*. British Journal of Sports Medicine, 2000. **34**(1): p. 6-6.
367. Koshoedo, S.A., et al., *Understanding the complex interplay of barriers to physical activity amongst black and minority ethnic groups in the United Kingdom: a qualitative synthesis using meta-ethnography*. BMC public health, 2015. **15**(1): p. 643.
368. Chinn, D.J., et al., *Barriers to physical activity and socioeconomic position: implications for health promotion*. Journal of Epidemiology and Community Health, 1999. **53**(3): p. 191.
369. Kjær, T., B. Højgaard, and D. Gyrd-Hansen, *Physical exercise versus shorter life expectancy? An investigation into preferences for physical activity using a stated preference approach*. Health Policy, 2019.
370. Koeneman, M.A., et al., *Determinants of physical activity and exercise in healthy older adults: a systematic review*. International Journal of Behavioral Nutrition and Physical Activity, 2011. **8**(1): p. 142.
371. Kirk, M.A. and R.E. Rhodes, *Occupation correlates of adults' participation in leisure-time physical activity: a systematic review*. American journal of preventive medicine, 2011. **40**(4): p. 476-485.
372. Orbell, S. and P. Sheeran, *'Inclined abstainers': A problem for predicting health-related behaviour*. British Journal of Social Psychology, 1998. **37**(2): p. 151-165.
373. Quicke, J.G., et al., *Relationship Between Attitudes and Beliefs and Physical Activity in Older Adults With Knee Pain: Secondary Analysis of a Randomized Controlled Trial*. Arthritis care & research, 2017. **69**(8): p. 1192-1200.
374. Mo, P.K., et al., *Physical Activity in People With Mental Illness in Hong Kong: Application of the Health Belief Model*. J Sport Exerc Psychol, 2016. **38**(2): p. 203-8.
375. Oja, P., et al., *Health benefits of different sport disciplines for adults: systematic review of observational and intervention studies with meta-analysis*. Br J Sports Med, 2015. **49**(7): p. 434-440.
376. Verhagen, E. and W. van Mechelen, *Health issues as primary reasons for choosing sport for all programs*. Promoting Sport for All Benefits and Strategies for the 21st Century, 2010: p. 43.
377. Guo, Y., et al., *Assessing the quality of mobile exercise apps based on the American College of Sports Medicine guidelines: a reliable and valid scoring instrument*. Journal of medical Internet research, 2017. **19**(3): p. e67.
378. Hagger, M.S., N.L. Chatzisarantis, and S.J. Biddle, *The influence of autonomous and controlling motives on physical activity intentions within the Theory of Planned Behaviour*. British journal of health psychology, 2002. **7**(3): p. 283-297.
379. Guttman, N. and C.T. Salmon, *Guilt, fear, stigma and knowledge gaps: ethical issues in public health communication interventions*. Bioethics, 2004. **18**(6): p. 531-552.
380. Conrad, P., *Wellness as virtue: Morality and the pursuit of health*. Culture, medicine and psychiatry, 1994. **18**(3): p. 385-401.
381. van Stralen, M.M., et al., *Determinants of initiation and maintenance of physical activity among older adults: a literature review*. Health Psychology Review, 2009. **3**(2): p. 147-207.
382. Trost, S.G., et al., *Correlates of adults' participation in physical activity: review and update*. Medicine & science in sports & exercise, 2002. **34**(12): p. 1996-2001.
383. Johnson, D., et al., *Gamification for health and wellbeing: a systematic review of the literature*. Internet Interv. 6: 89–106. 2016.



384. Haseler, C., R. Crooke, and T. Haseler, *Promoting physical activity to patients*. BMJ, 2019. **366**: p. l5230.
385. Saelens, B.E. and S.L. Handy, *Built environment correlates of walking: a review*. Medicine and science in sports and exercise, 2008. **40**(7 Suppl): p. S550.
386. Wendel-Vos, W., et al., *Potential environmental determinants of physical activity in adults: a systematic review*. Obesity reviews, 2007. **8**(5): p. 425-440.
387. Sawyer, A.D., et al., *Cross-sectional interactions between quality of the physical and social environment and self-reported physical activity in adults living in income-deprived communities*. PloS one, 2017. **12**(12): p. e0188962.
388. Beauchamp, M.R., et al., *Group-based physical activity for older adults (GOAL) randomized controlled trial: Exercise adherence outcomes*. Health Psychol, 2018. **37**(5): p. 451-461.
389. Richards, E.A. and Y. Cai, *Integrative Review of Nurse-delivered Community-Based Physical Activity Promotion*. Appl Nurs Res, 2016. **31**: p. 132-8.
390. Baretta, D., O. Perski, and P. Steca, *Exploring users' experiences of the uptake and adoption of physical activity apps: Longitudinal qualitative study*. JMIR mHealth and uHealth, 2019. **7**(2): p. e11636.
391. Di Domenico, S.I. and R.M. Ryan, *The Emerging Neuroscience of Intrinsic Motivation: A New Frontier in Self-Determination Research*. Front Hum Neurosci, 2017. **11**: p. 145.
392. Ormel, H.L., et al., *Predictors of adherence to exercise interventions during and after cancer treatment: A systematic review*. Psychooncology, 2018. **27**(3): p. 713-724.
393. Morrison, L., et al., *Optimizing engagement with Internet-based health behaviour change interventions: Comparison of self-assessment with and without tailored feedback using a mixed methods approach*. British journal of health psychology, 2014. **19**(4): p. 839-855.
394. Couper, M.P., et al., *Engagement and retention: measuring breadth and depth of participant use of an online intervention*. Journal of medical Internet research, 2010. **12**(4): p. e52.
395. Danaher, B.G., et al., *Defining participant exposure measures in Web-based health behavior change programs*. Journal of Medical Internet Research, 2006. **8**(3): p. e15.
396. Behavioural Insights Team. *EAST: Four simple ways to apply behavioural insights*. 2014; Available from: [https://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST\\_FA\\_WEB.pdf](https://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf).
397. Wachter, R., *Making IT work: harnessing the power of health information technology to improve care in England*. London, UK: Department of Health, 2016.
398. England, N., *Five year forward view*. 2014.
399. Association, L.G., *Using digital technology to improve the public's health: a guide for local authorities*. 2018.
400. Mayor of Greater Manchester. *Beelines: Greater Manchester's cycling and walking infrastructure proposal*. 2019; Available from: <https://assets.ctfassets.net/nv7y93idf4jq/6J11U1IT8kAq2oCISoEGYy/ae9e07dd0974abcfdb3434cdf46ab5b/Beelines.pdf>.
401. Sport England. *Call for sport and physical activity sector to embrace digital revolution*. 2019; Available from: <https://www.sportengland.org/news-and-features/news/2019/april/30/leaders-call-on-sport-and-physical-activity-sector-to-embrace-digital-revolution/>.
402. Kelders, S.M., et al., *Persuasive system design does matter: a systematic review of adherence to web-based interventions*. Journal of medical Internet research, 2012. **14**(6): p. e152.

403. Eysenbach, G., *The law of attrition*. Journal of medical Internet research, 2005. **7**(1): p. e11.
404. West, R., *Prime theory of motivation*. 2015.
405. West, R. and J. Brown, *Theory of addiction*. 2013: John Wiley & Sons.
406. Gawronski, B. and L.A. Creighton, *Dual-process theories*. The Oxford handbook of social cognition, 2013: p. 282-312.
407. Epstein, S., *Integration of the cognitive and the psychodynamic unconscious*. American psychologist, 1994. **49**(8): p. 709.
408. Kahneman, D., *Thinking, fast and slow*. 2011: Macmillan.
409. Strack, F. and R. Deutsch, *Reflective and impulsive determinants of social behavior*. Personality and social psychology review, 2004. **8**(3): p. 220-247.
410. Gawronski, B. and G.V. Bodenhausen, *Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change*. Psychological bulletin, 2006. **132**(5): p. 692.
411. Tversky, A. and D. Kahneman, *Judgment under uncertainty: Heuristics and biases*. Judgment and decision making: An interdisciplinary reader, 1986: p. 38-55.
412. Zenko, Z., P. Ekkekakis, and G. Kavetsos, *Changing minds: Bounded rationality and heuristic processes in exercise-related judgments and choices*. Sport, Exercise, and Performance Psychology, 2016. **5**(4): p. 337.
413. Williams, D.M. and L.C. Bohlen, *Motivation for exercise: Reflective desire versus hedonic dread*. 2019.
414. Dolan, P., G. Kavetsos, and I. Vlaev, *The happiness workout*. Social indicators research, 2014. **119**(3): p. 1363-1377.
415. Bauman, A., et al., *Is there sufficient evidence regarding signage-based stair use interventions? A sequential meta-analysis*. BMJ open, 2017. **7**(11): p. e012459.
416. Zimmerman, F.J., *Using behavioral economics to promote physical activity*. Preventive Medicine, 2009. **49**(4): p. 289-291.
417. Nocon, M., et al., *Increasing physical activity with point-of-choice prompts-a systematic review*. Scandinavian journal of public health, 2010. **38**(6): p. 633-638.
418. Hollands, G.J., et al., *The TIPPMIE intervention typology for changing environments to change behaviour*. Nature Human Behaviour, 2017. **1**(8): p. 0140.
419. Forberger, S., et al., *Nudging to move: a scoping review of the use of choice architecture interventions to promote physical activity in the general population*. International Journal of Behavioral Nutrition and Physical Activity, 2019. **16**(1): p. 77.
420. Courtin, E. and M. Knapp, *Social isolation, loneliness and health in old age: a scoping review*. Health Soc Care Community, 2017. **25**(3): p. 799-812.
421. Pearson, T., *Loneliness in adults: Evidence-based research and interventions for NPs*. The Nurse Practitioner, 2019. **44**(9): p. 26-34.
422. The Health Foundation, *A recipe for action: using wider evidence for a healthier UK*. 2018.
423. Povey, R., et al., *The theory of planned behaviour and healthy eating: Examining additive and moderating effects of social influence variables*. Psychology & Health, 2000. **14**(6): p. 991-1006.
424. Mermelstein, R., E. Lichtenstein, and K. McIntyre, *Partner support and relapse in smoking-cessation programs*. J Consult Clin Psychol, 1983. **51**(3): p. 465-6.
425. Caplan, S.E., *Relations among loneliness, social anxiety, and problematic Internet use*. CyberPsychology & behavior, 2006. **10**(2): p. 234-242.



426. Pittman, M. and B. Reich, *Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words*. Computers in Human Behavior, 2016. **62**: p. 155-167.
427. Requena, F. and L. Ayuso, *Individualism or complementarity? The effect of digital personal networks on face-to-face personal networks*. Information, Communication & Society, 2018: p. 1-15.
428. Coughlin, S.S., et al., *A Review of Smartphone Applications for Promoting Physical Activity*. Jacobs J Community Med, 2016. **2**(1).
429. McKay, F.H., et al., *Evaluating mobile phone applications for health behaviour change: a systematic review*. Journal of telemedicine and telecare, 2018. **24**(1): p. 22-30.
430. Bort-Roig, J., et al., *Measuring and influencing physical activity with smartphone technology: a systematic review*. Sports Med, 2014. **44**(5): p. 671-86.
431. Muellmann, S., et al., *Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review protocol*. Syst Rev, 2016. **5**: p. 47.
432. Cheatham, S.W., et al., *The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review*. J Sports Med Phys Fitness, 2018. **58**(4): p. 534-548.
433. Bock, C., M.N. Jarczok, and D. Litaker, *Community-based efforts to promote physical activity: a systematic review of interventions considering mode of delivery, study quality and population subgroups*. J Sci Med Sport, 2014. **17**(3): p. 276-82.
434. Jahangiry, L., et al., *Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials*. Public Health, 2017. **152**: p. 36-46.
435. Maon, S., et al., *The use of web-based interventions to prevent excessive weight gain*. J Telemed Telecare, 2012. **18**(1): p. 37-41.
436. Afshin, A., et al., *Information Technology and Lifestyle: A Systematic Evaluation of Internet and Mobile Interventions for Improving Diet, Physical Activity, Obesity, Tobacco, and Alcohol Use*. J Am Heart Assoc, 2016. **5**(9).
437. Stockwell, S., et al., *Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: A systematic review and meta-analysis*. Experimental gerontology, 2019.
438. Daeschner, M., *Empowering people, improving healthcare*. 2019.
439. Nahum-Shani, I., et al., *Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support*. Annals of Behavioral Medicine, 2017. **52**(6): p. 446-462.
440. Baretta, D., et al., *Implementation of the goal-setting components in popular physical activity apps: Review and content analysis*. Digital health, 2019. **5**: p. 2055207619862706.
441. Hardeman, W., et al., *A systematic review of just-in-time adaptive interventions (JITAs) to promote physical activity*. International Journal of Behavioral Nutrition and Physical Activity, 2019. **16**(1): p. 31.
442. Gill, S. and S. Panda, *A smartphone app reveals erratic diurnal eating patterns in humans that can be modulated for health benefits*. Cell metabolism, 2015. **22**(5): p. 789-798.
443. Chamberlain, K. and M. Murray, *Critical health psychology*. Critical psychology: An introduction, 2009. **2**: p. 144-158.
444. Lupton, D., *The digitally engaged patient: Self-monitoring and self-care in the digital health era*. Social Theory & Health, 2013. **11**(3): p. 256-270.
445. Tulloch, J. and D. Lupton, *Television, AIDS and risk: A cultural studies approach to health communication*. 1997.

446. Petersen, A. and D. Lupton, *The new public health: Health and self in the age of risk*. 1996: Sage Publications, Inc.
447. McGregor, S., *Neoliberalism and health care*. International Journal of Consumer Studies, 2001. **25**(2): p. 82-89.
448. Mielewczyk, F. and C. Willig, *Old clothes and an older look: The case for a radical makeover in health behaviour research*. Theory & Psychology, 2007. **17**(6): p. 811-837.
449. Lyons, A.C. and K. Chamberlain, *Health psychology: A critical introduction*. 2006: Cambridge University Press.
450. Piggitt, J. and L. Hart, *Physical activity advocacy in the UK: a multiple streams analysis of a hybrid policy issue*. Leisure Studies, 2017. **36**(5): p. 708-720.
451. Gustat, J., et al., *Lessons in promoting active living: the collaborative perspective*. J Public Health Manag Pract, 2013. **19**(3 Suppl 1): p. S58-64.
452. Yen, H.Y. and C. Li, *Determinants of physical activity: A path model based on an ecological model of active living*. PLoS One, 2019. **14**(7): p. e0220314.
453. Iwasaki, Y., C.P. Coyle, and J.W. Shank, *Leisure as a context for active living, recovery, health and life quality for persons with mental illness in a global context*. Health Promot Int, 2010. **25**(4): p. 483-94.
454. Sallis, J.F., et al., *An ecological approach to creating active living communities*. Annu Rev Public Health, 2006. **27**: p. 297-322.
455. Giovannucci, E., *A framework to understand diet, physical activity, body weight, and cancer risk*. Cancer Causes & Control, 2018. **29**(1): p. 1-6.
456. Wilmot, E.G., et al., *Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis*. Diabetologia, 2012. **55**(11): p. 2895-905.
457. Michie, S. and R. West, *A guide to development and evaluation of digital behaviour change interventions in healthcare*. 2016: Silverback Publishing.
458. Michie, S., L. Atkins, and R. West, *The behaviour change wheel: a guide to designing interventions*. . 1st ed. 2014, Great Britain: Silverback Publishing.
459. Johnston, M., et al., *Linking behavior change techniques and mechanisms of action: Triangulation of findings from literature synthesis and expert consensus*. 2018.
460. Craig, L.E., et al., *Identifying the barriers and enablers for a triage, treatment, and transfer clinical intervention to manage acute stroke patients in the emergency department: a systematic review using the theoretical domains framework (TDF)*. Implementation Science, 2016. **11**(1): p. 157.
461. Graham-Rowe, E., et al., *Barriers and enablers to diabetic retinopathy screening attendance: Protocol for a systematic review*. Systematic reviews, 2016. **5**(1): p. 134.
462. Laba, T.-L., et al., *Strategies to improve adherence to medications for cardiovascular diseases in socioeconomically disadvantaged populations: a systematic review*. International journal of cardiology, 2013. **167**(6): p. 2430-2440.





























































Appendix A: Apps used in the sample identification process (Study 1)

a) Health and Fitness: TOP PAID N=100 (all) DATE: 17/10/16 time 1:32pm

Top Paid iPhone Apps

Paid Free Top Grossing










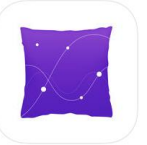




















Health & Fitness

 1. 7 Minute Workout... Health & Ftn... £2.29 In-Aoo Purchases	 2. The Wonder Weeks Health & Ftn... £1.49 In-Aoo Purchases	 3. Adrian James 6 Pack Abs... Health & Ftn... £0.79	 4. Full Fitness : Exercise... Health & Ftn... £2.29	 5. Instant Heart Rate+ Heart... Health & Ftn... £2.29 In-Aoo Purchases	 6. Mindfulness Daily Health & Ftn... £1.49 In-Aoo Purchases	 7. Yoga Studio Health & Ftn... £2.99	 8. Carbs & Cals - Count your... Health & Ftn... £3.99	 9. Couch to 5K@ - Runnin... Health & Ftn... £1.49	 10. Runtastic PRO Running... Health & Ftn... £3.99 In-Aoo Purchases	 11. CWP & Me Health & Ftn... £0.99	 12. Adrian James High... Health & Ftn... £0.79
 13. Ultimate Food Value... Health & Ftn... £2.99 In-Aoo Purchases	 14. Running for Weight Loss... Health & Ftn... £1.49 In-Aoo Purchases	 15. Nois!l Health & Ftn... £1.49	 16. KetoDiet Basic Health & Ftn... £1.49 In-Aoo Purchases	 17. Period Tracker Deluxe Health & Ftn... £1.49	 18. Instant Fitness : 600+... Health & Ftn... £2.29 In-Aoo Purchases	 19. Rainy Mood - Rain Sounds &... Health & Ftn... £2.29	 20. buddhify - mindfulness &... Health & Ftn... £3.99	 21. Bodyweight Training: You... Health & Ftn... In-Aoo Purchases	 22. Virtual Gastric Band... Health & Ftn... £4.99	 23. Push ups 0 to 100: push u... Health & Ftn... £1.49	 24. PAUSE - Relaxation at... Health & Ftn... £1.49
 25. CarbsControl - ... Health & Ftn... £2.29 In-Aoo Purchases	 26. Couch to 5K PRO Health & Ftn... £2.29	 27. My Macros+ Diet. Weight... Health & Ftn... £2.29 In-Aoo Purchases	 28. mySymptoms... Health & Ftn... £2.29	 29. Body Weight Fat loss... Health & Ftn... £0.79	 30. InnerFire Health & Ftn... £2.99	 31. Footsteps - Pedometer Health & Ftn... £2.29	 32. Points Calculator &... Health & Ftn... £2.29	 33. Hormone Horoscope Pro Health & Ftn... £1.49	 34. Moodnotes - Thought... Health & Ftn... £2.99	 35. Muscle 2 - iPhone Edition Health & Ftn... £2.29	 36. Great Abs: Sit ups... Health & Ftn... £2.29
 37. Police Fitness - Bleep... Health & Ftn... £0.79	 38. Runtastic Pull Ups PRO... Health & Ftn... £1.99	 39. StopWatch + Timer and... Health & Ftn... £0.79	 40. Placebo - Reprogramm... Health & Ftn... £4.99	 41. Vedic Med Health & Ftn... £2.99	 42. ManageMyFat... Health & Ftn... £3.99	 43. How to Eat a Balanced Di... Health & Ftn... £2.49	 44. The Work App Health & Ftn... £0.79	 45. Fast Exercise Health & Ftn... £1.49	 46. White Noise with Andrew... Health & Ftn... £0.79 In-Aoo Purchases	 47. Deep Sleep with Andrew... Health & Ftn... £2.29	 48. Model My Diet - Women... Health & Ftn... £1.49
 49. Half Marathon Health & Ftn... £4.49	 50. HeartWatch. Heart & Activ... Health & Ftn... £2.29	 51. Map My Ride+ - GPS... Health & Ftn... £2.29 In-Aoo Purchases	 52. Healthy Weight Loss... 5 Apps £7.99	 53. LDNM: The Workout App Health & Ftn... £1.49 In-Aoo Purchases	 54. Good Morning Alar... Health & Ftn... £2.99	 55. Paleo (to) - The Ultimate... Health & Ftn... £0.79	 56. Life Cycle - Chart your life Health & Ftn... £1.49	 57. Adrian James... Health & Ftn... £0.79	 58. Stamina - Longer lasting... Health & Ftn... In-Aoo Purchases	 59. MindFit Health & Ftn... £2.99	 60. 5K to 10K Health & Ftn... £2.29





b) iTunes: Health and Fitness: TOP FREE N=100 (all) DATE: 19/10/16 time: 10:30

 <p>1. Calorie Counter &amp; Diet... Health &amp; Fitness + Update In-App Purchases</p>	 <p>2. Fitbit Health &amp; Fitness Download In-App Purchases</p>	 <p>3. Headspace: Guided... Health &amp; Fitness Get In-App Purchases</p>	 <p>4. Sweat with Kayla - Fitness... Health &amp; Fitness + Get In-App Purchases</p>	 <p>5. Running for Weight Loss:... Health &amp; Fitness Get In-App Purchases</p>	 <p>6. Sleep Cycle alarm clock Health &amp; Fitness + Download In-App Purchases</p>	 <p>7. Map My Run - GPS Running &amp;... Health &amp; Fitness Download In-App Purchases</p>	 <p>8. Clue: Period, PMS, fertility &amp;... Health &amp; Fitness Get In-App Purchases</p>	 <p>9. Strava Running and... Health &amp; Fitness Download In-App Purchases</p>	 <p>10. Pillow: The sleep cycle... Health &amp; Fitness + Get In-App Purchases</p>
 <p>11. Pacer - Pedometer plu... Health &amp; Fitness Download In-App Purchases</p>	 <p>12. Walking for Weight Loss:... Health &amp; Fitness Get In-App Purchases</p>	 <p>13. Asana Rebel - Yoga for... Health &amp; Fitness + Get In-App Purchases</p>	 <p>14. my PureGym Health &amp; Fitness Get In-App Purchases</p>	 <p>15. Period Diary (Period &amp;... Health &amp; Fitness Get In-App Purchases</p>	 <p>16. Steps Pedometer &amp;... Health &amp; Fitness Download In-App Purchases</p>	 <p>17. FitStar Personal Trainer Health &amp; Fitness + Get In-App Purchases</p>	 <p>18. Period Tracker Lite Health &amp; Fitness Get In-App Purchases</p>	 <p>19. David Lloyd Clubs Health &amp; Fitness Get In-App Purchases</p>	 <p>20. Superdrug Health &amp; Fitness Get In-App Purchases</p>
 <p>21. Instant Heart Rate: Heart Rat... Health &amp; Fitness Get In-App Purchases</p>	 <p>22. Weight Watchers Health &amp; Fitness + Get In-App Purchases</p>	 <p>23. Seven - 7 Minute Workou... Health &amp; Fitness + Download In-App Purchases</p>	 <p>24. Couch to 5K Runner, 0 to 5... Health &amp; Fitness + Update In-App Purchases</p>	 <p>25. Good Morning Alarm... Health &amp; Fitness + Get In-App Purchases</p>	 <p>26. My Pregnancy &amp;... Health &amp; Fitness + Get In-App Purchases</p>	 <p>27. Moves Health &amp; Fitness Download In-App Purchases</p>	 <p>28. Lose It! - Weight Loss... Health &amp; Fitness + Get In-App Purchases</p>	 <p>29. Workout: Gym personal... Health &amp; Fitness Get In-App Purchases</p>	 <p>30. Runkeeper - GPS Running Health &amp; Fitness Get In-App Purchases</p>



31. Nike+ Run Club  
Health & Fitness

Download



32. Pedometer++  
Health & Fitness

Get

In-App Purchases



33. Garmin Connect™...  
Health & Fitness

Download



34. Calm: Meditation...  
Health & Fitness

Get

In-App Purchases



35. Relax Melodies: Slee...  
Health & Fitness

Get

In-App Purchases



36. Calorie Counter +  
Health & Fitness

+ Get

In-App Purchases



37. 30 Day Ab Challenge FREE  
Health & Fitness

Get

In-App Purchases



38. Stepz - Pedometer &...  
Health & Fitness

Get

In-App Purchases



39. Period Tracker - My...  
Health & Fitness

Get

In-App Purchases



40. JOHNSON'S®...  
Health & Fitness

Get



41. Interval Timer - Timing...  
Health & Fitness

+ Get



42. SworKit - Custom...  
Health & Fitness

+ Get

In-App Purchases



43. Daily Workouts FREE...  
Health & Fitness

+ Download



44. Runtastic GPS Running,...  
Health & Fitness

Download

In-App Purchases



45. Moment - Screen Time...  
Health & Fitness

+ Get

In-App Purchases



46. Runtastic Results: Body...  
Health & Fitness

+ Get

In-App Purchases



47. Health Mate - Steps tracker...  
Health & Fitness

Get

In-App Purchases



48. 100% Army Fit  
Health & Fitness

Get



49. My Diet Coach - ...  
Health & Fitness

+ Get

In-App Purchases



50. Patient Access  
Health & Fitness

+ Get



51. Exercise4Less Fitness Partner  
Health & Fitness

Get



52. Smoke Free - Quit smoking...  
Health & Fitness

Get

In-App Purchases



53. Running, Walking and...  
Health & Fitness

+ Get

In-App Purchases



54. Ab & Core SworKit - Free...  
Health & Fitness

+ Get

In-App Purchases



55. Nike+ Training Club  
Health & Fitness

Get



56. Lifesum - Healthier eatin...  
Health & Fitness

+ Get

In-App Purchases



57. Record by Under Armour,...  
Health & Fitness

Get



58. 7 Minutes Workout - ...  
Health & Fitness

+ Get

In-App Purchases



59. Life - Period Tracker,...  
Health & Fitness

Get

In-App Purchases



60. Digipill - Sleep,...  
Health & Fitness

+ Get

In-App Purchases



61. BabyBump  
Pregnancy Pro...  
Health & Fitness

+ Get ▾



62. Seconds -  
Interval Timer  
Health & Fitness

+ Get ▾

In-App Purchases



63. SleepBot -  
Smart Cycle...  
Health & Fitness

Get ▾



64. One You  
Couch to 5K  
Health & Fitness

Download ▾



65. Sound  
Sleeper: white...  
Health & Fitness

+ Get ▾

In-App Purchases



66. 30 Day  
Fitness...  
Health & Fitness

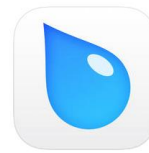
+ Get ▾



67. Map My Ride  
- GPS Cycling ...  
Health & Fitness

Get ▾

In-App Purchases



68. Rain Rain  
Sleep Sounds  
Health & Fitness

+ Get ▾

In-App Purchases



69. Everyone  
Active  
Health & Fitness

Get ▾



70. Vibrating  
Massager FREE  
Health & Fitness

Get ▾



71. Glow -  
Fertility Tracker...  
Health & Fitness

Update ▾

In-App Purchases



72. Pregnancy &  
Baby | What to...  
Health & Fitness

+ Get ▾



73. SkinVision -  
Melanoma skin...  
Health & Fitness

Get ▾

In-App Purchases



74. MyNetDiary -  
Calorie/KJ...  
Health & Fitness

Get ▾

In-App Purchases



75. Cycles -  
Period tracker...  
Health & Fitness

+ Get ▾

In-App Purchases



76. St John  
Ambulance  
FIRST AID  
Health & Fitness

Get ▾



77. White Noise  
Free: sounds f...  
Health & Fitness

+ Get ▾



78. Daily Yoga -  
Workouts, ...  
Health & Fitness

+ Get ▾

In-App Purchases



79. The  
Mindfulness...  
Health & Fitness

Get ▾

In-App Purchases



80. Freeletics  
Bodyweight - ...  
Health & Fitness

+ Get ▾

In-App Purchases



81. Baby's  
heartbeat Free  
Health & Fitness

Get ▾



82. Fitness  
Buddy: Gym &...  
Health & Fitness

Get ▾

In-App Purchases



83. Breast  
Check Now  
Health & Fitness

Get ▾



84. 30 Day  
Squat Challenge...  
Health & Fitness

Get ▾

In-App Purchases



85. Fertility  
Friend -...  
Health & Fitness

+ Get ▾

In-App Purchases



86. MINDBODY -  
Book Fitness...  
Health & Fitness

Get ▾



87. Rthm - The  
Life App  
Health & Fitness

Get ▾



88. Sleep Well  
Hypnosis -...  
Health & Fitness

+ Get ▾

In-App Purchases



89. Skyfit: On-  
Demand Audio...  
Health & Fitness

Get ▾

In-App Purchases



















































90. Sleep  
Matters  
Health & Fitness











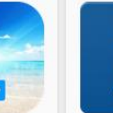








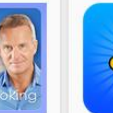









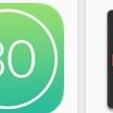
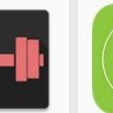













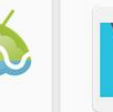


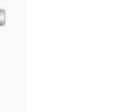
Get ▾

























c) Google Play: Health and Fitness: TOP PAID N=100 DATE: 17/10/16 time 1:45pm





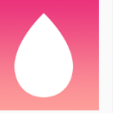
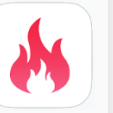

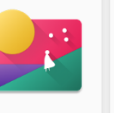

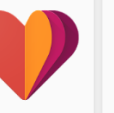
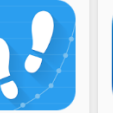
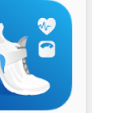






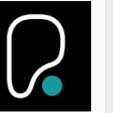


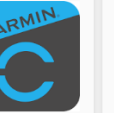






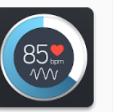





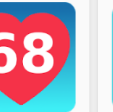
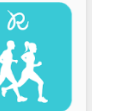








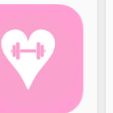


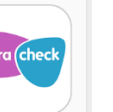






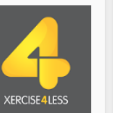

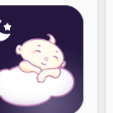
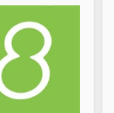
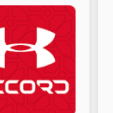
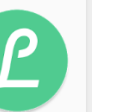
 1. Sleep Cycle alarm Northcube AS ★★★★★ £0.99	 2. The Wonder Weir Domus Technica ★★★★★ £1.29	 3. 7 Minute Workout Fitness Guide Inc. ★★★★★ £2.48	 4. Carbs & Cals - Diet Chello Publishing Limited ★★★★★ £3.99	 5. Tools & Mi Band Zdeněk Horák ★★★★★ £1.89	 6. buddhify - mindfulness Mindfulness Everywhere ★★★★★ £1.99	 7. Runtastic PRO Running Runtastic ★★★★★ £4.99	 8. Couch to 5K® ACTIVE Network, LLC ★★★★★ £1.67	 9. Virtual Gastric Band The Happy Apps Company ★★★★★ £4.99	 10. Period Tracker GP International LLC ★★★★★ £1.24	 11. Deep Sleep with HiveBrain Software ★★★★★ £1.89	 12. CWP & Me Cambridge Manufacturer ★★★★★ £0.99
 13. Yoga Studio Gaiam ★★★★★ £3.19	 14. Fitness Buddy Azumio, Inc. ★★★★★ £1.89	 15. Rainy Mood Plain Theory, Inc. ★★★★★ £2.79	 16. White Noise TMSOFT ★★★★★ £0.89	 17. Ultimate Food Fenlander Software Solutions ★★★★★ £2.49	 18. C25K® - 5K Running Zen Labs Fitness ★★★★★ £1.59	 19. The Mindfulness MindApps ★★★★★ £1.49	 20. Jason's 5-Day Juice Juice Master ★★★★★ £4.99	 21. My Diet Coach InspiredApps (A.L) LTD ★★★★★ £2.99	 22. Jason's Super Food Juice Master ★★★★★ £5.49	 23. Map My Ride+ MapMyFitness, Inc. ★★★★★ £2.30	 24. Mindfit Mindfit ★★★★★ £2.99
 25. CRON-O-Meter Cronometer Inc. ★★★★★ £1.65	 26. Fitness Trainer Your personal coach ★★★★★ £0.99	 27. Home workout Arusotech ★★★★★ £2.99	 28. Just 6 Weeks Heavyraid ★★★★★ £0.99	 29. Terri Ann's 123 Terri-Ann 123 Diet Plan ★★★★★ £2.98	 30. Get Running Splendid Things ★★★★★ £2.39	 31. Zombies, Run! Six to Start ★★★★★ £1.79	 32. Gastric Band Slim Positive ★★★★★ £3.60	 33. Twilight Pro Urbandroid Team ★★★★★ £2.99	 34. KetoDiet Basic Compumaster Ltd ★★★★★ £1.49	 35. Map My Walk+ MapMyFitness, Inc. ★★★★★ £2.52	 36. iMuscle 2 3D4Medical.com, LLC ★★★★★ £1.89
 37. Point Tracker PointTracker ★★★★★ £1.00	 38. 5:2 Diet Complete Time Inc. (UK) Ltd ★★★★★ £1.99	 39. I Can Make You Once Byten Ltd ★★★★★ £4.99	 40. Run with Map MapMyFitness, Inc. ★★★★★ £2.30	 41. Sleep Lullaby Urbandroid Team ★★★★★ £1.99	 42. LDNM: The World Puzzle London ★★★★★ £1.79	 43. 10K Running Zen Labs Fitness ★★★★★ £2.49	 44. Seconds Pro Runloop ★★★★★ £3.99	 45. Paleo (io) The Impossible X LLC ★★★★★ £0.78	 46. Simply Being Meditation Oasis ★★★★★ £1.55	 47. Beep Test Pro Ruval Enterprises ★★★★★ £0.60	 48. Cook & Count Health Apps Ltd ★★★★★ £3.99






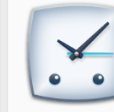


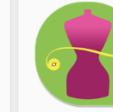
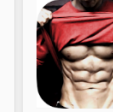
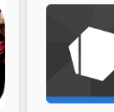




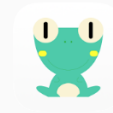


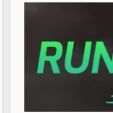
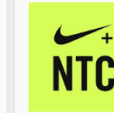


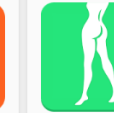
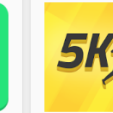



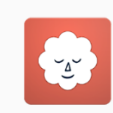



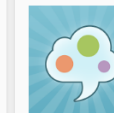

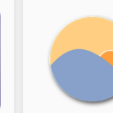

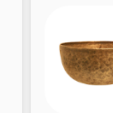





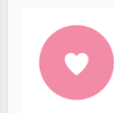
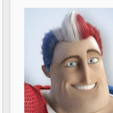


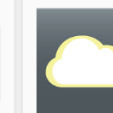
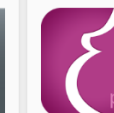
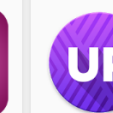


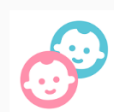
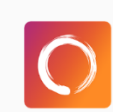

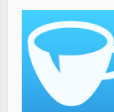


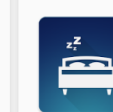
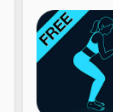

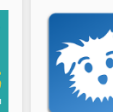


 49. Medical ID - ICE Laurent Pellegrino ★★★★★ £1.69	 50. Chloe Madeley Hungrydog Media Ltd ★★★★★ £2.99	 51. Relax Melodies Ipnos Software ★★★★★ £2.29	 52. mySymptoms F SkyGazer Labs Ltd ★★★★★ £2.29	 53. Baby Shusher Baby Shusher LLC ★★★★★ £3.00	 54. Wendler 531 Pro SaraSoft ★★★★★ £2.99	 55. I Can Make You Once Byten Ltd ★★★★★ £4.99	 56. InnerFire Innerfire ★★★★★ £3.11	 57. 5K to 10K ACTIVE Network, LLC ★★★★★ £2.32	 58. Marathon Training Zen Labs Fitness ★★★★★ £8.49	 59. Relax VR: Rest Now VR Now VR ★★★★★ £1.79	 60. MapMyFitness+ MapMyFitness, Inc. ★★★★★ £2.30
 61. Runtastic Push-Pro Runtastic ★★★★★ £1.99	 62. Jason's 3-Day Juice Master Juice Master ★★★★★ £4.49	 63. 7 Minute Workout Pro Simple Design Ltd. ★★★★★ £2.39	 64. Virtual Trainer Virtual Trainer ★★★★★ £3.49	 65. Starting Strength Shabu Pty Ltd ★★★★★ £8.99	 66. C9 Tracker RHLS Software Ltd ★★★★★ £0.99	 67. Instant Heart Rate Azumio Inc. ★★★★★ £2.99	 68. Stop Smoking Help Diviniti Publishing Ltd ★★★★★ £3.99	 69. Lose Weight - A HiveBrain Software ★★★★★ £1.90	 70. Tabata Pro - Tabata SimpleTouch LLC ★★★★★ £2.50	 71. Gentle Wakeup Dr. Alexander Rieger ★★★★★ £3.09	 72. Mindfulness: The Mindful Creation ★★★★★ £2.52
 73. WomanLog Pro Pro Active App ★★★★★ £2.56	 74. PAUSE ustwo ★★★★★ £1.49	 75. You Are Your Own Leafcutter Studios ★★★★★ £3.99	 76. Fit Journal Pro Sultan Seidalin ★★★★★ £1.59	 77. Fast Tract Diet Norm Robillard, Ph.D. ★★★★★ £6.37	 78. 30 Day Fitness Foamy Media Ltd. ★★★★★ £2.39	 79. Symmetric Strength Symmetric Strength ★★★★★ £3.59	 80. Half Marathon Training Zen Labs Fitness ★★★★★ £8.49	 81. myWOD - #1 Workout MetaBot, LLC ★★★★★ £3.08	 82. The Blood Type 4 North American Pharm ★★★★★ £2.42	 83. Gymster - Gym Gymster ★★★★★ £0.59	 84. Perfect Body Builder Bodybuilding-Apps.com ★★★★★ £2.59
 85. Runtastic Heart Pro Runtastic ★★★★★ £1.99	 86. The Fast Metabolism Diet Penguin Random House ★★★★★ £2.45	 87. Hydro Coach Pro Oodium App Ideas ★★★★★ £2.49	 88. PDC Pole Dance Community PDC Publishing ★★★★★ £2.99	 89. MMA Spartan Workouts Arsutech ★★★★★ £2.99	 90. Jason Vale's 5:2 Juice Diet Juice Master ★★★★★ £4.99	 91. Runtastic Mountain Runtastic ★★★★★ £4.99	 92. Jason's 7-Day Juice Juice Master ★★★★★ £5.99	 93. Mindroid PRO Urbandroid Team ★★★★★ £2.99	 94. Bluelight Filter Hardy-infinity ★★★★★ £2.99	 95. Runtastic Road Runtastic ★★★★★ £4.99	 96. A HIIT License Pimpim Mobile ★★★★★ £0.89

 <p>97. TravelBit Premium Steven Shewfelt</p> <p>★★★★★ £2.09</p>	 <p>98. The Walk: Fitness Six to Start</p> <p>★★★★★ £2.59</p>	 <p>99. Serenity Simple Celerity Software</p> <p>★★★★★ £0.93</p>	 <p>100. Modern Essence AromaTools</p> <p>★★★★★ £5.56</p>	 <p>101. Mindfulness mindspace</p> <p>★★★★★ £1.50</p>	 <p>102. The Plan Tracker Leightons Logic</p> <p>★★★★★ £1.50</p>	 <p>103. FibroMapp Bodymap Apps</p> <p>★★★★★ £2.39</p>	 <p>104. Visualize Health HiveBrain Software</p> <p>★★★★★ £1.89</p>	 <p>105. Relax: Stress &amp; Anxiety Relief Saagara</p> <p>★★★★★ £1.99</p>	 <p>106. SyncMyTracks SyncMyTracks</p> <p>★★★★★ £2.00</p>	 <p>107. MyDays Pro Christian Albert Mueller</p> <p>★★★★★ £1.44</p>	 <p>108. 21 Day Countdown Cube Software Solution</p> <p>★★★★★ £1.49</p>
 <p>109. Tai Chi 13 Postures YMAA</p> <p>★★★★★ £0.78</p>	 <p>110. Weigh-In Deluxe Wil Corp. Software</p> <p>★★★★★ £1.28</p>	 <p>111. Your Positive Mindfulness Coach The Happy Apps Comp</p> <p>★★★★★ £4.99</p>	 <p>112. Fitness Point Zero One GmbH</p> <p>★★★★★ £4.69</p>	 <p>113. Vegan Alternatives AYUSH</p> <p>★★★★★ £0.69</p>	 <p>114. EasyLoss The Happy Apps Comp</p> <p>★★★★★ £4.99</p>	 <p>115. Wear Stand-Up DYNA Logix</p> <p>★★★★★ £1.59</p>	 <p>116. Mike Rashid Grand Apps</p> <p>★★★★★ £1.21</p>	 <p>117. Runtastic Pedometer Runtastic</p> <p>★★★★★ £1.99</p>	 <p>118. iTrackBites Swellapps Inc.</p> <p>★★★★★ £2.99</p>	 <p>119. White Noise TMSOFT</p> <p>★★★★★ £2.69</p>	 <p>120. HIIT - Interval Caynax</p> <p>★★★★★ £0.99</p>

d) Google Play: Health and Fitness: TOP FREE N=100 DATE: 18/10/16 time 21: 56pm

 1. GO Clock - Alarm GO Dev Team ★★★★★	 2. Fitbit Fitbit, Inc. ★★★★★	 3. Calorie Counter MyFitnessPal, Inc. ★★★★★	 4. Headspace - Meditation Headspace, Inc. ★★★★★	 5. Sweat With Kayla Sweat with Kayla ★★★★★	 6. Running for Wellbeing Red Rock Apps ★★★★★	 7. Strava Running Strava Inc. ★★★★★	 8. Fabulous - Motivation TheFabulous ★★★★★	 9. Daily Yoga JeremiahHardint ★★★★★	 10. Google Fit - Fitness Google Inc. ★★★★★	 11. Pedometer tayutau ★★★★★	 12. Pedometer & Pacer Pacer Health ★★★★★
 13. 30 Day Fit Challenge NorthPark.Android ★★★★★	 14. adidas train & run adidas ★★★★★	 15. Period Tracker Simple Design Ltd ★★★★★	 16. Runtastic Run Runtastic ★★★★★	 17. Walk with Map MapMyFitness, Inc. ★★★★★	 18. Clue - Period Tracker BioWink GmbH ★★★★★	 19. my PureGym Pure Gym Ltd ★★★★★	 20. Twilight Urbandroid Team ★★★★★	 21. Weight Watchers Weight Watchers ★★★★★	 22. Garmin Connect Garmin ★★★★★	 23. Run with Map MapMyFitness, Inc. ★★★★★	 24. Superdrug Superdrug Stores ★★★★★
 25. Flo Period & Contraception OWHEALTH, INC. ★★★★★	 26. S Health Samsung Electronics ★★★★★	 27. Calm - Meditation Calm.com, Inc. ★★★★★	 28. 7 Minute Workout Simple Design Ltd ★★★★★	 29. Instant Heart Rate Azumio Inc. ★★★★★	 30. Moves ProtoGeo ★★★★★	 31. Patient Access Patient.info ★★★★★	 32. Skyfit - Audio Skyfit Sports ★★★★★	 33. Period Tracker GP International ★★★★★	 34. Geocaching Groundspeak Inc ★★★★★	 35. Unique Heart Rate Meet Your Need ★★★★★	 36. Runkeeper - GPS FitnessKeeper, Inc. ★★★★★
 37. ViewRanger Augmentra ★★★★★	 38. Nike+ Run Club Nike, Inc. ★★★★★	 39. Home workout Arsutech ★★★★★	 40. One You Couch Public Health England ★★★★★	 41. Lose It! FitNow, Inc. ★★★★★	 42. Smoke Free David Crane ★★★★★	 43. Water Drink NorthPark.Android ★★★★★	 44. St John Ambulance St John Ambulance ★★★★★	 45. Fit Girls Guide Fit Girls ★★★★★	 46. Coach by Cigna Cigna ★★★★★	 47. Change4Life Public Health England ★★★★★	 48. Calorie Count Nutracheck ★★★★★
 49. Map My Ride MapMyFitness, Inc. ★★★★★	 50. Gym Workout Fitness22 ★★★★★	 51. Weight Loss aktiWir GmbH ★★★★★	 52. Free Pedometer 4Free Studio ★★★★★	 53. Fitness & Body VGFIT LLC ★★★★★	 54. 100% Army Fitness ARMY JOBS UK ★★★★★	 55. Xercise4Less Virtuaqym Profesion ★★★★★	 56. C25K Zen Labs Fitness ★★★★★	 57. JOHNSON'S GiantSky ★★★★★	 58. 8fit - Workout 8fit ★★★★★	 59. Under Armour Under Armour ★★★★★	 60. Lifesum - The Lifesum ★★★★★

											
61. Relax Melodie Ipnos Software ★★★★★	62. Endomondo - Endomondo.com ★★★★★	63. 30 Day Ab Challenge Viquor Apps ★★★★★	64. Interval Timer dreamspark ★★★★★	65. 5K Run - Couch Vandersoft ★★★★★	66. SleepBot - Sleep SleepBot ★★★★★	67. Monitor Your Husain Al-Bustar ★★★★★	68. SnoreLab: Re- Reviva Software ★★★★★	69. My Diet Coach InspiredApps (A.I ★★★★★	70. 6 Pack Promise ATHLEANX.COM ★★★★★	71. Freeletics Body Freeletics ★★★★★	72. Veryfit 2.0 Smart Wearable ★★★★★
											
73. White Noise B AMICOOLSOFT ★★★★★	74. Oral-B App Procter & Gambl ★★★★★	75. Everyone Active Innovatise UG ★★★★★	76. Zeroner Health com.kunekt ★★★★★	77. Nature Sounds Relaxio ★★★★★	78. Bounts Bounts Limited ★★★★★	79. Running Distanc Fitness22 ★★★★★	80. Nike+ Training Nike, Inc. ★★★★★	81. Couch to 10K Zen Labs Fitness ★★★★★	82. Mi Fit Xiaomi Technolo ★★★★★	83. Butts, legs and ShvaqerFM ★★★★★	84. 5K Runner: 0 Fitness22 ★★★★★
											
85. FitNotes - Gyr James Gay ★★★★★	86. White Noise F TMSOFT ★★★★★	87. Slimming Traine Worrall Studios L ★★★★★	88. Stop Breathe Tools for Peace ★★★★★	89. Pregnancy by Bounty (UK) Ltd ★★★★★	90. Buddhist Medit Spacebuq ★★★★★	91. First aid by Br British Red Cross ★★★★★	92. Self-help Anxi University of the ★★★★★	93. Pregnancy Tr Everyday Health ★★★★★	94. f.lux (preview) justqetflux ★★★★★	95. BlueLight Filter Hardy-infinity ★★★★★	96. Insight Timer Insight Network I ★★★★★
											
97. Belly Fat Expert Expert Lab Team ★★★★★	98. StrongLifts 5x5 StrongLifts ★★★★★	99. Tabata Timer Evgeniy Sharafar ★★★★★	100. Calorie Counter SparkPeople ★★★★★	101. Map My Fitness MapMyFitness, Inc ★★★★★	102. Breast Check Breast Cancer N ★★★★★	103. Aquafresh B GSK ★★★★★	104. Fitatu Calorie Fitatu ★★★★★	105. Push Ups W NorthPark.Andro ★★★★★	106. mywellness Mywellness srl ★★★★★	107. BabyBump F Alt12 Apps, Inc. ★★★★★	108. UP@ - Smart Jawbone ★★★★★
											
109. Hydro Coach Codium App Idea ★★★★★	110. Brain in Hand Brain in Hand ★★★★★	111. Baby Names Health & Parentin ★★★★★	112. MINDBODY MINDBODY Inc ★★★★★	113. HPlus Watch HPlus ★★★★★	114. 7 Cups - Any 7 Cups of Tea ★★★★★	115. TomTom My TomTom Interna ★★★★★	116. Sworik Personal Nexercise Apps, I ★★★★★	117. Sleep Better Runtastic ★★★★★	118. 30 Day Butt Viquor Apps ★★★★★	119. One You Easy Public Health Enc ★★★★★	120. Down Dog: C Yoga Buddhi Co. ★★★★★



## Appendix B: Screenshot of the data extraction form (Study 1)

Duplicate (available in both stores)		Duplicate - SCC	Duplicate
<b>DESCRIPTIVE DATA</b>			
App's Name	Fitbit	Strava Running and Cycling GPS	
1 Rank			
2 Brief description	tracker (mainly walking and running)		
3 Type of PA targeted (e.g., running, walking, yoga, whole body)	predominantly running (accelerometer-based)		
4 Platform on which the app is available (iTunes, GP, Both)	Both	Both	
5 Developer's name	Fitbit, Inc	Strava, Inc.	
<b>iTunes</b>			
1 Cost	£ -	£ -	
2 Size (in megabites)	104	92.3	
3 Last update	09/12/2016	09/01/2017	
4 Version	2.3	5.10.0	
5 Number of ratings	7871	10742	
6 Weighted mean rating (weighted by nr of ratings in iTunes)	3.9	4.6	
7 Store weight (proportion of total reviews from itunes)	4%	5%	
<b>Google Play</b>			
1 Cost	£ -	£ -	
2 Size (in megabites)	22.35	21.41	
3 Last update	15/12/2016	03/01/2017	
4 Version	2.4	5.10.0	
5 Number of ratings	207000	199596	
6 Weighted mean rating (weighted by nr of ratings in GP)	4	4.5	
7 Store weight (proportion of total reviews from GP)	96%	95%	
<b>Development / Evaluation of the app</b>			
1 Organisational Affiliation *1	Commercial	Commercial	
Is there a reference to an expert being involved in the app development/evaluation (check app description/the app's website, the app itself)? Indicate who or "No" *2	No	No	
3 Were potential users involved in the development of an app?	No	No	
4 Are there any studies in peer-reviewed journals associated with the app? *3 (put the links to the studies in the sheet: peer-reviewed studies)	Yes	Yes	
5 What other apps have they created? How many of these health related ( / / )	0	0/0	
<b>Data safety and security (keep record of the link to the privacy policy to rationalise t</b>			
1 Is there privacy information available? Only continue if answered 'Yes'	Yes	Yes	
2 Is the privacy information available without the need to download the app? (check: app store via a link to the privacy policy or the app's website)	Yes	Yes	
3 Is the privacy information available within the app?	Yes	Yes	
4 Is there a short form notice (in plain English) highlighting key data practices which are disclosed in detail in the full privacy policy? (ONLY indicate N/A if the policy is already short)	Yes	Yes	
5 Is the privacy policy available in any other languages?	No	No	
6 Does the app collect personally identifiable information?	Yes	Yes	
7 Does the app share users' data with 3rd party?	Yes	Yes	
8 Does the app say how the users' data security is ensured? e.g. encryption, authentication, firewall system	No	Yes	
<b>System Usability Scale</b>			
1 I think that I would like to use this app frequently	Agree	Disagree	
2 I found the app unnecessarily complex.	Disagree	Strongly agree	
3 I thought the app was easy to use.	Neutral	Disagree	
4 I think that I would need the support of a technical person to be able to use this app	Disagree	Neutral	
5 I found the various functions in this app were well integrated.	Neutral	Neutral	
6 I thought there was too much inconsistency in this app.	Strongly disagree	Neutral	
7 I would imagine that most people would learn to use this app very quickly.	Disagree	Strongly disagree	
8 I found the app very awkward to use.	Disagree	Neutral	
9 I felt very confident using the app.	Agree	Disagree	
10 I needed to learn a lot of things before I could get going with this app.	Disagree	Agree	
<b>1. Goals and planning</b>			
1.1. Goal setting (behavior)	Present ++	Absent	
1.2. Problem solving	Absent	Absent	
1.3. Goal setting (outcome)	Present ++	Absent	
1.4. Action planning	Absent	Absent	
1.5. Review behavior goal(s)	Absent	Absent	
1.6. Discrepancy between current behavior and goal	Absent	Absent	
1.7. Review outcome goal(s)	Absent	Absent	
1.8. Behavioral contract	Absent	Absent	
1.9 Commitment	Absent	Absent	

## **Manual for the data privacy and security assessment**

Purpose of this quality indicator is to assess whether there are privacy information available to the user (or potential user), how accessible the information is, and whether the statement covers basic information about data gathering.

Rationale for inclusion of these questions included in the assessment:

Information Commissioner's Office (ICO), Online Trust Alliance (OTA), and the Global Privacy Enforcement Network (GPEN) recommendations:

- Links must be included to privacy policies before, during and after downloading the app (GPEN)
- Use plain English (ICO) <sup>[1]</sup><sub>SEP</sub>
- Use language appropriate to your audience (ICO)
- Transparency about purpose is crucial. Developers should say which data you want, say why (ICO)
- While the app may be in English, having the privacy policy and terms of use in other languages is highly recommended to maximize user's ability in comprehending the app's data practices (OTA0).
- A short form notice highlighting key data practices which are disclosed in detail in the full privacy policy (especially as users might access the statement on a small screen of their phone) should be considered (OTA)

There are various types of documents that pertain to the security and privacy of the data. LOOK FOR:” Privacy policy,” “Terms of service” or “Legal”, etc.

Pasted from the Extraction Form (do not complete this).

Data privacy and security assessment		
Q1: Availability	Is there privacy information available? (only continue if answered 'Yes')	Yes/No
Q2: Availability	Is the privacy information available without the need to download the app? (example: app store, via a link to the privacy policy or the app's website)	Yes/No
Q3: Availability	Is the privacy information available within the app?	Yes/No
Q4: Accessibility	Is there a short form notice (in plain English) highlighting key data practices which are disclosed in detail in the full privacy policy? (ONLY indicate NA if the policy is already short (see example below)	Yes/No/NA
Q5: Accessibility	Is the privacy policy available in any other languages?	Yes/No
Q6: Data gathering	Does the app collect personally identifiable information?	Yes/No/NS
Q7: Data sharing	Does the app share users' data with 3rd party?	Yes/No/NS
Q8: Data security	Does the app say how the users' data security is ensured? e.g. encryption, authentication, firewall system	Yes/No

The questions pertain to:

Q1: Availability
Q2: Availability
Q3: Availability
Q4: Accessibility
Q5: Accessibility
Q6: Data gathering
Q7: Data sharing
Q8: Data security

Instructions on how to complete the table in the Extraction Form

Q1	Availability	Indicate	Where to find this out?	What to look for?
	Is there privacy information available for the user (or potential user)?	YES/NO	<ul style="list-style-type: none"> <li>• app description in app store</li> <li>• via a link to the privacy policy</li> <li>• the app's website (if exists)</li> <li>• in the app itself</li> </ul>	the words "Privacy policy," "Terms of service" or "Legal."
NOTE: Only if answered "YES" to the Q1 proceed to the assessment below				

Please download or copy the privacy statement if possible or copy the link to it as we might need to assess it. Note that it is up to you where you record it. It might be needed when we compare results.

	Location: Is the privacy information available in:	Indicate	Notes
Q2	Is the privacy information available in app store, via a link to the privacy policy or the app's website?	YES/NO	Q2 from Q3 needs to be distinguished as having the policy only within the app is not the best practice
Q3	Is the privacy information available within the app?	YES/NO	
Q4	Is there a short form notice (in plain English) highlighting key data practices which are disclosed in detail in the full privacy policy? (ONLY indicate N/A if the policy is already short)	YES/NO/NA	Best practice
Q5	Is the privacy policy and terms of use available in any other languages?	YES/NO	
Q6	Does the app collect personally identifiable information?	YES/NO/NS	
Q7	Does the app share users' data with 3 <sup>rd</sup> party?	YES/NO/NS	
Q8	Does the app say how the users' data security is ensured? e.g. encryption, authentication, firewall system	YES/NO	



## Notes about extraction of specific questions

Regarding Q6: Does the app collect personally identifiable information?

Personal information' means any details about a living individual that can be used on its own, or with other data, to identify the user.

Personally identifiable information (PII) can also be called sensitive personal information (this is a subset of PII). In the documents we will review it can also say just "personal information" or "sensitive information".

Regarding Q4: Is there a short form notice (in plain English) highlighting key data practices which are disclosed in detail in the full privacy policy?

Q4	Is there a short form notice (in plain English) highlighting key data practices which are disclosed in detail in the full privacy policy? (ONLY indicate NS if the policy is already short (see example below)	Yes/No/NA
----	--	-----------

### SOME EXAMPLES to aid the extraction

#### 1) Very vague statement

Below is an example of a very vague privacy policy. It is not really transparent, not concrete. ICO says:

"Transparency about purpose is crucial. Don't just say which data you want, say why."

In this case you would answer 'NA' to question 4 as the statement is short. However, please note that the policy actually doesn't say anything, i.e., does not address what it does with the personal information, does it share it with anyone, how it protects it, etc.

Therefore, the answer to other questions in the assessment might be 'NA' as well. It is not enough that the study mentions data. It happens that a developer can create a generic policy online (there are companies that do so) but you can see that this one is very vague.

## Privacy Policy

Your privacy is important to us, and it is Quick Fit's policy to respect your privacy regarding any information we may collect while operating our app. Accordingly, we have developed this Policy in order for you to understand how we collect, use, communicate and disclose and make use of personal information. The following outlines our privacy policy.

- Before or at the time of collecting personal information, we will identify the purposes for which information is being collected.
- We will collect and use of personal information solely with the objective of fulfilling those purposes specified by us and for other compatible purposes, unless we obtain the consent of the individual concerned or as required by law.
- We will only retain personal information as long as necessary for the fulfillment of those purposes.
- We will collect personal information by lawful and fair means and, where appropriate, with the knowledge or consent of the individual concerned.
- Personal data should be relevant to the purposes for which it is to be used, and, to the extent necessary for those purposes, should be accurate, complete, and up-to-date.
- We will protect personal information by reasonable security safeguards against loss or theft, as well as unauthorized access, disclosure, copying, use or modification.
- We will make readily available to customers information about our policies and practices relating to the management of personal information.

We are committed to conducting our business in accordance with these principles in order to ensure that the confidentiality of personal information is protected and maintained. Quick Fit may change its Privacy Policy from time to time, and at Quick Fit's sole discretion.

2) Although the statement below is short, it is informative as they do not collect any identifiable information – therefore answer 'NA' to Q4 would be appropriate

## Privacy Policy

---

Last Updated: April 27, 2015

Palm Shadow Apps LLC (also referred to as "we", "us", "our") value your privacy. This privacy statement describes what data we may collect and for what purpose.

### Personally Identifiable Information

Palm Shadow Apps LLC does not collect any personally identifiable information.

### Non-identifying Information

Palm Shadow Apps LLC may collect some non-identifying information such as the name and version of operating system, etc.

We use this information for analytical purposes to improve our apps and services.

We do NOT share this information with third parties.

### Changes to this Policy

We may change this Policy from time to time to reflect changes in our apps and services so please review this policy periodically. When we post changes, we will revise the last updated date of the policy.

Appendix D: SUS descriptive record for both reviewers (Study 1)

a) Reviewer 1 (PB): SUS descriptive record

App's name	Overall usability of the app	Issues
Strava Running and Cycling GPS	a lot of options. Two sets of settings make it confusing, hard to find features - like history of the PA. Not easy to learn and memorise. Navigation difficult - Hard to get back to some of the features I accessed before. Good quality graphics	Errors: unable to see summary of activity as it says that there is an error - my data may have been deleted on Strava server. Cannot set goal in the free version (says no goal set which is discouraging).
Pacer – Pedometer plus Weight Loss and BMI Tracker (called Pedometer & Weight Loss Coach in GP)	A lot of features and it seems a bit hard to get back to certain options- not easy to memorise when the features are. Menu bar is not always available. Takes a while to learn and memorise but great features (although quite a large variety) High quality graphic, pleasant to use. Great visualisation of the data. Moderately efficient.	commercials pop up straight away. Complex but worth it.
Map My Run – GPS Running & Workout Tracker	The app has a lot of functions but is well organised. Navigation is easy after some exploration. Good menu. Quite memorable and learnable. Also, quite efficient although it take a while to log the exercise.	commercials get in a way a bit
Adidas train & run	takes a while to get used to as there are many features. Not quite sure what the different colour coding means for the goal It would have been helpful if there was a tutorial at the beginning. Otherwise quite pleasant to use, some good features. High quality graphics, good menu.	it is not clear when they want the user to get a sensor device. Confusing a bit if the app is enough to use all the features. It is not
Steps Pedometer & Step Counter Activity Tracker	Good tutorial at the beginning. Simple but pleasant and quality design. efficient, memorable, learnable. No errors.	no tailoring, at least the goal could be tailored
Home workout MMA Spartan Free	Easy to navigate, straight forward, good menu facilitates navigation. Easy to learn, memorable and efficient. Not many commercials in the free version - advantage	Crushed when tried to do the workout. Quite uninteresting graphics, basic. Blatantly uses stereotype of an overly masculine man.

Stepz – Pedometer & Step Counter for Tracking Steps	Great menu bar on the bottom that stays there means user can navigate easily and efficiently. Easy to use, great beautiful design, efficient, memorable. No errors. Seems to be very little commercials, very subtle way to give the option to upgrade.	
Interval Timer – Timing for HIIT Training and Workouts	high quality graphics. Easy to use, learn and memorise.	would be better is the social media sharing would include what has been done in terms of the exercise
Sworkit – Custom Workouts for Exercise & Fitness	easy to use, high quality design, good menu bar, efficient, memorable. Takes just a bit of time to learn as there are a lot of workouts to choose from. No errors.	
Fitness & Bodybuilding	It is easy to navigate and to learn to use the app. Memorable and quite efficient. Would be better if the app could record the work put automatically. It would have been more efficient. No errors. Standard graphics (high quality) quite boring white screen graphics.	Slightly difficult to record the exercise: the screen flickers when trying to put the weights in. commercials popping up slightly annoying but not too much in comparison to other apps.
30 Day Ab Challenge FREE	Simple app. Basic graphics. I was confused about how to start a challenge and how to log the exercise. There was no indication and it was not intuitive. Took me about 5-7 minutes to find out how to do this. Once learnt that, the app is memorable and efficient.	Commercials get a bit in the way.
Runtastic Results: Body Workout Fitness Trainer	easy to use, high quality design although needs to DOWNLOAD the videos, good menu bar, efficient, memorable. No errors.	
C25K® - 5K Running Trainer	take a bit to learn (not too much) because the buttons are not very intuitive. Pleasant to use (great graphics). user can accomplish tasks quite efficiently once learnt the buttons. Quite easy to remember.	commercial get slightly in the way of the training
Health Mate – Steps tracker & Life coach by Withings	Efficient, quite easy to learn once you learn to navigate the app. Needs a bit of exploration. Great menu bar on the bottom that stays there means user can navigate easily and efficiently. Quite Easy to use considering	Not sure how to find - heart rate monitor through camera - it's advertised but does not show how to.

	that there are a lot of features, great beautiful design. No errors.	
Running, Walking and Biking with Endomondo (called Endomondo...in GP)	The app has a lot of functions and it take a while to explore these. Navigation is easy after some exploration. Good menu. Moderately memorable and learnable. Also, quite efficient. Standard quality design.	No errors.
Map My Ride – GPS Cycling & Route Tracker	The app has a lot of functions but is well organised. Navigation is easy after some exploration. Good menu. Quite memorable and learnable. Also, quite efficient although it takes a while to log the exercise.	commercials get in a way a bit
Interval Timer	easy to use but only one function - a timer. Memorable, learnable, no errors but very basic. Good graphics	Would be great if user could save the timers as to view the progress
5K Run - Couch to 5K	Quite boring graphics; app is memorable and easy to learn. Efficient.	Commercials get in the way a bit; Quite basic, no tailoring at all for the user.
7 Minutes Workout – Women Fitness Exercise Trainer	very easy to use, high quality design, good menu bar, efficient, memorable.	cannot set the weight measurement metrics - only in lbs. commercials get in the way
Seconds – Interval Timer	Takes a bit to learn to navigate considering that this is a timer. Takes a while to understand how to start the timer. Moderately memorable and learnable. Good options for workout timers in-built. A lot of functions to learn in order to be efficient with the timer. Some of these functions seem unnecessary. The useful options of recording the time is not available in the free version.	starts with the add to buy the full version. Many functions, setting weight, maximum volume of oxygen (VO2) but not sure what is the use of it.
Running Distance Tracker +	Easy to use app, memorable	could be perceived as complex for some

Freeletics Bodyweight – Workout and Training	Quite easy to use once you get through the features. When doing the exercise, it would be helpful to have the audio as cannot see any explanation, only the user would have to get back to the videos (and disrupt their workout). high quality design, (although have to DOWNLOAD the videos) good menu bar, efficient, memorable.	Error: "cannot save training at the moment"
Couch to 10K Running Trainer	take a bit to learn (not too much) because the buttons are not very intuitive. Pleasant to use (great graphics). user can accomplish tasks quite efficiently once learnt the buttons. Quite easy to remember.	commercial get slightly in the way of the training
FitNotes - Gym Workout Log	Takes time to understand the structure of the app. Takes a while to memorise and learn the app. Not so efficient, a bit confusing how to set up a workout set. 1RM is an advantage, not seen anywhere else. Graphics are basic, like a log - no visuals.	cannot see the menu bar on the side, it's too far on the left.
Belly Fat Exercises	simple app, only information and photos. Graphics standard. Not easy to memorise, a lot of writing. Takes time to learn the app.	a lot of written information; commercials pop quite a bit. Some corrections needed in the written material.
Belly Fat Workout FREE – 10 Minute Ab Exercises	simple app but not so easy to navigate at first as the menu is not easily available so sometimes get stuck (low efficiency). Otherwise memorable and learnable.	hard to navigate, length of exercise not written (just small icons); below average graphics. Lots of commercials, also attempting to get users to buy the subscription
Movesum – Step counter by Lifesum	Good tutorial at the beginning. Simple but pleasant and quality design. efficient, memorable, learnable. No errors.	
Adrian James 6 Pack Abs Workout	Easy to use, very simple design, memorable and easily learnt. Good professional graphics	The only annoying this is that there is a commercial as the starting page although the app is paid.
Full Fitness: Exercise Workout Trainer	Quite complex. Confusing when adding the exercise to the workout routine. Over-complicated. This makes it less memorable and less learnable. Standard to high quality graphics. No errors. All in all,	unable to create a workout as I am not sure how to add exercise. Too many different methods to add exercise and workout's. Would be better if there was a timer to time the whole workout

	very difficult to learn. Advantage: great categorisation of the exercises based on muscle group.	
Runtastic PRO Running and Workout Tracker (Runtastic PRO Running, Fitness in GP)	Quite a lot of functions, menu bar full. Otherwise the app has great graphics (pleasant to use), is easy to navigate. High learnability, user can accomplish tasks quite efficiently. Quite easy to remember. No errors. Simplicity is a great advantage.	a lot of functions of the menu bar.
Adrian James High Intensity Interval Training	Easy to use, very simple design, memorable and easily learnt. Good professional graphics	The only annoying this is that there is a commercial as the starting page although the app is paid.
Running for Weight Loss PRO: training plan, GPS, how-to-lose-weight tips	Quite a lot of information, especially written information. Otherwise the app has great graphics (pleasant to use), is easy to navigate. High learnability, user can accomplish tasks quite efficiently. Quite easy to remember. No errors	adds popping and easy to click on them by mistake when using the app although the apps was not free.
Fitness Trainer FULL version	Good, simple design. Memorable and easy to learn. Good quality graphics. No errors.	Would be better if there was a timer to time the whole workout
Instant Fitness : 600+ exercises, 100+ workouts...	There are a lot of exercises but they are arranged quite well. Standard-low graphics, do not look highly pleasing. Easy to learn and memorise. Great menu bar on the bottom that stays there means user can navigate easily and efficiently.	Errors: the app got stuck one an exercise while the timer was still running. Low quality photo when posted on twitter.
Push ups 0 to 100: push ups challenge trainer pro	very easy to use, great beautiful design, good menu bar, efficient, memorable. No errors.	heavy on promoting other apps of the same developer. Bit annoying
Couch to 5K Runner, 0 to 5K run training	very easy to use, great beautiful design, good menu bar, efficient, memorable. No errors.	heavy on promoting other apps of the same developer. Bit annoying
iMuscle 2 – iPhone Edition	Advantage: tips while using the app; also, a tour on how to use the app but not straight forward where it is as it is hidden in the settings. Great visualisation of the muscles! good graphics although black background is hard to look at for a long time. Quite complex but I feel it is learnable after a bit of time, also	Difficult moving around, the buttons are quite small and do not work sometimes cannot read the description if the workout is being done, no explanation of different types of exercises. At the beginning a bit confusing about how to start or add workouts. Also, cannot find work out without the

	efficient one the user gets used to it.	equipment, just exercises. Menu a bit awkward to use.
10K Running Trainer Pro	take a bit to learn (not too much) because the buttons are not very intuitive. Pleasant to use (great graphics). user can accomplish tasks quite efficiently once learnt the buttons. Quite easy to remember. No errors	not intuitive menu. A bit hard to navigate at first. Buttons not too intuitive
Police Fitness – Bleep Test	easy to use, learnable, memorable, very basic.	could not download the second test. Very basic, could have some trainings in there to increase the results.
Marathon Trainer - 26.2 42K	take a bit to learn (not too much) because the buttons are not very intuitive. Pleasant to use (great graphics). user can accomplish tasks quite efficiently once learnt the buttons. Quite easy to remember. No errors	not intuitive menu. A bit hard to navigate at first. Buttons not too intuitive
MapMyFitness + Workout Trainer	a lot of features but these are well integrated. Not so easy to memorise as there are a lot of options but easy to learn and efficient. Good quality graphics.	as you open the app when signing up there was no option to put my year in 1983 and I had to tap it so many times to get to 1983
5K to 10K	Easy to navigate. Great menu bar at the bottom means that I can access the main functions easily. Quite easily learnable and memorable although a lot of information. No errors; pleasant and professional graphics.	feels there is a lot of information at times but the pop up instructions of features help
7 Minute Workout Pro	Basic design, not very consistent graphics (high quality photos with primitive robot-like audio and basic graphics) and confusing forum. Simple to use afterwards when I get used to it. Quite memorable.	inconsistency in graphics, demonstration of behaviour - sends the user to you tube, a bit annoying.
10K Pacer: Run pace training. Run faster	very easy to use, great beautiful design, good menu bar, efficient, memorable. No errors.	heavy on promoting other apps of the same developer. Bit annoying
Starting Strength Official	Quite a lot of information, especially written information. Otherwise the app has great graphics (pleasant to use), is easy to navigate. High	A lot of written information



	learnability, user can accomplish tasks quite efficiently. Quite easy to remember. No errors	
Thor Fitness: 60 Day Bodyweight Workout Routine	Simple app, easy to learn, efficient. Memorable. Graphics not very high quality but basic and easy.	What is missing is the ability to look at the exercise beforehand, to know what to expect. error when posting on face book. When posting on twitter it says (null) Only option to have the phone in horizontal way
Half Marathon Trainer 13.1 21K	take a bit to learn (not too much) because the buttons are not very intuitive. Pleasant to use (great graphics). user can accomplish tasks quite efficiently once learnt the buttons. Quite easy to remember. No errors	not intuitive menu. A bit hard to navigate at first. Buttons not too intuitive
PDC Pole Dance Syllabus	quite easy to use but a bit primitive. Learnable, efficient, memorable. No errors. Basic graphics	
MMA Spartan Workouts Pro	Easy to navigate, straight forward, good menu facilitates navigation. Easy to learn, memorable and efficient.	Quite uninteresting graphics, basic. Blatantly uses stereotype of a overly masculine man.
Get Running (Coach to 5K)	Simplicity is a great advantage. Really great but simple graphics. Great visualisation of the plan of 9 weeks of running. High efficiency although 1 section (Talk) not intuitive and not explained. Otherwise, app highly memorable and pleasant to use. No errors overall but the app lacks consistency of design	Some sections are not explained - e.g., status updates on social media
WalkJogRun GPS Running Routes	The app takes a while to navigate around it. There is very little guidance on the various features. It takes a while to get used to it. The design is quite poor and the app is not very pleasant to use.	Crushed twice: when tried to create a route; when turned on settings. Also, unable to scroll down and tap on the bottom lines as the menu covers it.
Runtastic Road Bike PRO	Simple to use. The side menu that pops from the side is not very good. I was not sure what is in each section - not easily memorable. Otherwise high-quality design and intuitive. Can perform tasks quite quickly once you get used to the app.	with signing up there was no option to type in the birth date so the user has to scroll through each month since 2008!

Runtastic Mountain Bike PRO GPS Biking Computer, Trail and Route Tracker	quite a simple app to use. Good menu bar on the botton, makes navigation easier. Good and high-quality design. Fairly pleasant to use. No errors. Can perform tasks quite quickly once you get used to the app.	no issues
Chloe Madeley 15-minute fat loss workout	Very simple app with a great demo of the exercise and great graphics. Very simple but no way to tailor the apps in any way or turn off the ticking timer (which is a bit annoying). Simplicity isa big plus. See below, major disadvantage.	very difficult to memorise the exercise. When the timer starts there is no visual reminder of how the exercise look like.
CARROT Fit – 7 Minute Workout, Step Counter & Weight Tracker	Very easy to learn, efficient, memorable. Great simple but high-quality design.	I have some numbers (like unread emails) notifications as shown on the app logo but I cannot see what they were. Finally, I realised these were the steps taken today.

b) Reviewer 2 (GA): SUS descriptive record

App's name	Overall usability of the app	Issues
Strava Running and Cycling GPS	The app is easy to use but the fact that almost everything needs a premium membership makes it a bit awkward.	
Pacer – Pedometer plus Weight Loss and BMI Tracker (called Pedometer & Weight Loss Coach in GP)	The app has an an attractive interface but it recorded the steps even when it was the first time to open it without registering. It has a lot of functions, which might make it a bit confusing. The premium options are misleading.	
Map My Run – GPS Running & Workout Tracker	Having ads in the app was unpleasant and distracting. The app seems to have lots of options for the same function which can be confusing.	
Adidas train & run	The app has a nice interface with options to change the background but the graphs are confusing and not very clear. It is a bit confusing to move from the me screen to the training or running ones (i.e. what is the difference between them).	

Steps Pedometer & Step Counter Activity Tracker	A very simple interface. The app was working in the background, which can drain battery and there was no notification of that. The screen doesn't show that if you slide up, you will find the history of your steps table. The app although simple but confusing and not very clear.	
Home workout MMA Spartan Free	Simple and clear interface. Popup ads are distracting. Between the exercises-rest time- is confusing since it shows a message that it is time to rest but the background image shows someone exercising.	
Stepz – Pedometer & Step Counter for Tracking Steps	Simple and clear interface. Some annoying ads in the bottom.	
Interval Timer – Timing for HIIT Training and Workouts	Very simple interface with nice colours. The ad is below, it doesn't pop out so less distracting. There colours change during the different intervals.	
Sworki – Custom Workouts for Exercise & Fitness	The app is deceiving in not showing that plans for exercise are not available for free until the user chooses the plan that fits him/her. The app interface us attractive and clear. It has more than one language but that does not include privacy policy.	
Fitness & Bodybuilding	The app had apps in the bottom but were not distracting, however sometimes apps pop out as full screen. It required downloading videos, so that can be annoying if there is no internet. The workout plans videos do not work continuously, you need to press them.	
30 Day Ab Challenge FREE	A very simple and clear app.	
Runtastic Results: Body Workout Fitness Trainer	The app has an attractive and simple interface. it is easy to use but most of the important	

	functions needs to be paid for to access it.	
C25K® - 5K Running Trainer	The app is very simple and has an attractive interface. However, the functions are limited.	
Health Mate – Steps tracker & Life coach by Withings	The app has a very attractive interface and is easy to use.	
Running, Walking and Biking with Endomondo (called Endomondo...in GP)	The app has an attractive interface and lots usable free content. The setting menu has lots of functions to allow users to control the content.	
Map My Ride – GPS Cycling & Route Tracker	The interface is simple and attractive but the app seems to have lots of options for the same functions, which can be confusing.	
Interval Timer	The app is very simple. Going from the timer screen to the main screen is not simple, and you need to start a new timer each time you use the app.	
5K Run - Couch to 5K	Very simple and easy app but with not lots of options and doesn't measure steps or distance correctly.	
7 Minutes Workout – Women Fitness Exercise Trainer	simple and easy app but ads at the bottom are distracting.	
Interval Timer	The app was just a screen with a timer. The screen colours were very bright and unattractive but the displayed numbers were very clear.	
Running Distance Tracker +	Easy to use app with an attractive interface and interesting functions that work smoothly.	
Freeletics Bodyweight – Workout and Training	Attractive interface and you start by tailoring the app to your needs and current PA state. You have to confirm sign up with an email, which can be annoying. The videos of the workout have a sign that indicate they can be downloaded but they don't download. The videos for the exercises can be downloaded.	

	Also, the timer is not clear for which type of exercise.	
Couch to 10K Running Trainer	The app is very simple and has an attractive interface. However, pop up Ads are distracting.	
FitNotes - Gym Workout Log	The app functions are not clear and it is not an intuitive app to use.	
Belly Fat Exercises	The app has an unattractive interface with lots of text and stock pictures.	
Belly Fat Workout FREE – 10 Minute Ab Exercises	It has video ads which are extremely annoying and distracting. The interface was not very attractive but the videos were clear. Most of the apps functions were locked. The app starts by saying a free a trial for 7 days or upgrade by paying money but in reality, some functions are free for more than 7 days.	
Movesum – Step counter by Lifesum	The app has a very attractive interface and is simple to use.	
Adrian James 6 Pack Abs Workout	The app has an attractive interface and is simple to use. It has many helpful functions.	
Full Fitness: Exercise Workout Trainer	The app has an attractive interface albeit the colours are a little bit dull]. It has many helpful functions and exercises. The user can customise exercises based on the target muscle. However, logging data in is required for lots of things which can make it tedious, but helpful to personalise it to users.	
Runtastic PRO Running and Workout Tracker (Runtastic PRO Running, Fitness in GP)	The app has an attractive and simple interface. it is easy to use; some functions need more money to pay for.	
Adrian James High Intensity Interval Training	The app is very simple to use and has an attractive interface. If a workout is	

	chosen, users will not be able to switch between exercises.	
Running for Weight Loss PRO: training plan, GPS, how-to-lose-weight tips	the app has a simple and very attractive interface.	
Fitness Trainer FULL version	The app has many functions such as timer, stop watch, exercises, progress/tracker but they are not very well-integrated, the user has to enter the information himself. The information button made using the app easier.	
Instant Fitness: 600+ exercises, 100+ workouts...	The app is easy to use and has an attractive and simple interface. The function for sharing through social media was not working except in the calendar section.	
Push ups 0 to 100: push ups challenge trainer pro	A very simple apps and extremely attractive interface.	
Couch to 5K Runner, 0 to 5K run training	A very simple apps and extremely attractive interface.	
iMuscle 2 – iPhone Edition	The app was complicated and moving between the menu and the main screen was not very smooth. The app does have many, detailed functions.	
10K Running Trainer Pro	The app is very simple and has an attractive interface.	
Police Fitness – Bleep Test	The app is very simple but can be confusing for those not aware of the police physical assessment tests.	
Marathon Trainer - 26.2 42K	The app is very simple and has an attractive interface.	
MapMyFitness+ Workout Trainer	The interface is simple and attractive but the app seems to have lots of options for the same functions, which can be confusing.	
5K to 10K	The app is very simple and has an attractive interface.	
7 Minute Workout Pro	the app is very simple and easy to use. It has an attractive interface. The videos for the app are from you tube, so it might make the app lag a little bit.	

10K Pacer: Run pace training. Run faster	The app is simple to use and has an attractive interface.	
Starting Strength Official	The app is very complex especially the specific part about doing the exercise. It also contains the book with lots of text which can be annoying and time consuming. The videos are not attractive at all and the graph of progress is not clear.	
Thor Fitness: 60 Day Bodyweight Workout Routine	the app start with a video (very loud), which can be annoying. The app stays on the flipped mode for screening and doesn't change which is extremely annoying. The home page does not work. The timing in the app is not working properly. The work out pause was not working.	This app was not working properly at all. I might have missed some BCTs as every time I chose some workouts or programs, some won't work but those that did, were exceeding more than their time and not finishing at all, e.g a workout for 30 minutes did not finish even after more than an hour.
Half Marathon Trainer 13.1 21K	The app is very simple and has an attractive interface.	
PDC Pole Dance Syllabus	Simple app to view exercises but not many functions or instructions. The app is very huge, which is very inconvenient.	
MMA Spartan Workouts Pro	Simple and clear interface. Between the exercises-rest time- is confusing since it shows a message that it is time to rest but the background image shows someone exercising.	
Get Running (Coach to 5K)	the app is very simple to use, has large fonts and has a fun and unique interface.	
WalkJogRun GPS Running Routes	The app is simple but it can be confusing where to start running and tracking. It shows an image that sensors via Bluetooth will not be available except in the new version because the older one was crashing.	
Runtastic Road Bike PRO	The app is very simple to use and has an attractive interface.	

Runtastic Mountain Bike PRO GPS Biking Computer, Trail and Route Tracker	The app is very simple to use and has an attractive interface. It contains lots of easy function but it allows lots of apps, sesnors and others to connect to which can be confusing to users.	
Chloe Madeley 15-minute fat loss workout	The app is very easy to use and has an attractive interface. There is not lots of functions but its simplicity in use and design makes it usable and attractive.	
CARROT Fit – 7 Minute Workout, Step Counter & Weight Tracker	The app is simple to use and has a good interface. The voice of the coach was annoying and unpleasant.	*I really disliked this app, it was not funny, it was annoying and degrading.
Chloe Madeley Weights 4 Women	very simple app and simple to use. Main menu bar always visible. Easy to navigate and easy to learn and memorise. Pleasant, high quality design	
One You Couch to 5K	quite easy to use. Just a bit confusing with choosing the active days but. Would be better on a calendar. Navigation is good. No button to go to main menu directly but shows a button to choose an option at any time	
Runtastic Running & Fitness	It's a complex app with some great functions. Not so easy to learn. Sometimes found it hard to get to certain functions. Lots of different buttons. Overall, pleasant to use. No errors.	
Couch to 5K® - Running App and Training Coach	Easy to navigate. Good menu bar at the bottom means that I can access the main functions easily. Quite easily learnable and memorable although a lot of information. No errors; pleasant and professional graphics.	
Fitbit	Great menu button at the bottom, enables easy navigation between different features of the app. Quite complex, takes a while to get learn the app. Easy to forget the setup. No errors. Professional and pleasant design.	



Yoga Break	easy to use. Pleasant, quite simple design. Easy to learn and memorise. No errors although some descriptions of postures are not available.	
7 Minute Workout	Basic design, not very consistent graphics (high quality photos with primitive robot-like audio and basic graphics) and confusing forum. Plus: simple to use once you get past the ads.	
7 Minute Workout Challenge by Fitness Guide Inc	very easy to use, intuitive	
Footsteps – Pedometer	a bit confusing at some points, complicated with all the different measures. Not sure how to start a run or walk (a session)	
Daily Workouts FREE	Not sure what some buttons do - there are 2 buttons that are not self-explanatory. Graphics are not great but this is minor. Overall, easy app to use	

## **BCT Extraction Manual (V6) for physical activity apps**

This Manual is based on:

- initial notes on the extraction of BCTs
- testing the extraction of the BCTs between two reviewers on 5 apps (piloting)
- the BCTs online training
- the specific advice of a BCT expert (Lou Atkins)

### **Why the Manual?**

I realised it was important to have such a Manual as learning principle 2 (below) could not be directly applied to the extraction of BCTs from app as it is sometimes necessary that features of apps need to be 'translated' into BCTs. The Manual attempts to make it clearer when the BCT is present and when it is not present.

### **Level of confidence**

The level of confidence will be selected regarding the identification whereby:

+ BCT present in all probability

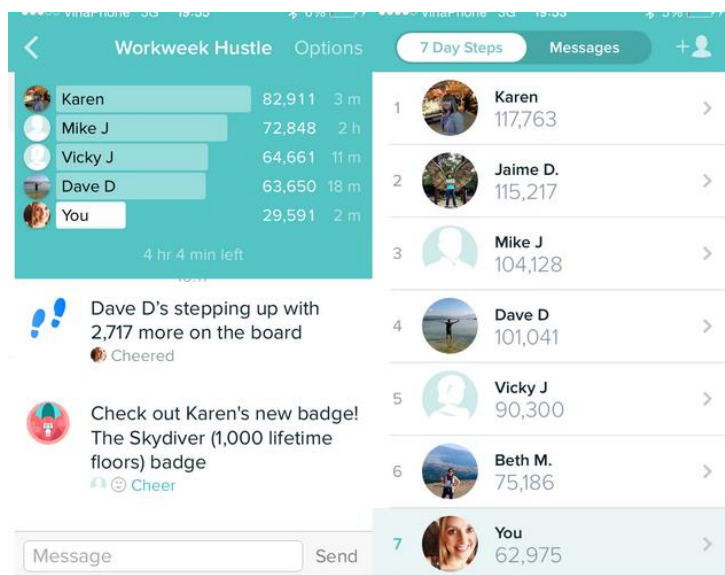
++ BCT present beyond all reasonable doubt

Please, do not think of it as - if you are not sure then put "+". It's purely about the strength of the evidence that the BCT is there.

The example below illustrated the "+". It is an example from your data extraction. I think it's a great example as you were not sure as well whether one of the BCTs is present. It was the 6.2 Social Comparison BCT in One You in Success Stories section.

11/20/2017

This would be a weaker evidence "+" than being a part of a group where daily physical activities are compared and ranked. You can see this in the example below (source: Fitbit) "++". It's more explicit that members of the group compare themselves as their results are presented. I also confirmed that with Lou Atkins.



## How to record if an app has 2 features corresponding to 1 BCT

If an app has, for example, a stopwatch to time exercise duration, that would be an example of self-monitoring of the behaviour. The app also might recommend users to keep a diary of when they exercise, also self-monitoring the behaviour but delivered in a different way. You would say the technique was used but delivered in two different ways.

Please note this in the extraction in the evidence for presence of BCTs, For example:

App's Features	PRESENCE of BCTs	Evidence
Timer Diary (free text)	2.3 Self-monitoring of behaviour	1) stopwatch; 2) diary

## What to do if the feature does not work?

If there seems to be a feature of but it does not work – put it in the feature's list and add - DOES NOT WORK. The example was the Prompts in the Results app.

## Notes of BCTs online training

### The 4 Learning Principles (from the online training):

1. Only code BCTs that are directly applied to the target behaviour (s) and population(s). DO NOT code a BCT unless the text explicitly links it to both (1) the named target behaviour(s) and (2) the named population(s).
2. Do not infer the presence of a BCT. The description must correspond to the definition of the BCT given in the taxonomy. If you are unsure, DO NOT code the BCT as present (NOTE: this principle might not apply to app extraction strictly. It may be less clear when extracting data from the apps than to extract the BCTs from the written material).
3. Take care distinguishing between BCTs that only differ in terms of their behaviour change type (i.e. behaviour vs. outcome).
4. Code technical terms and packages of BCTs that map onto BCTs in the taxonomy

In relation to the Principle 1 code the physical activity (PA) apps only in relation to the target behaviour - PA:

In the PA app assessment if the app is asking the user to record their calorie intake then this will not be a BCT to extract as it does not directly relate to the target behaviour, i.e., PA.

### 9.1 Credible Source

9.1	<i>Credible source</i>	Present verbal or visual communication from a credible source in favour of or against the behavior <i>Note: code this BCT if source generally agreed on as credible e.g., health professionals, celebrities or words used to indicate expertise or leader in field and if the communication has the aim of persuading; if information about health consequences, also code 5.1, Information about health consequences, if about emotional consequences, also code 5.6, Information about emotional consequences; if about social, environmental or unspecified consequences also code 5.3, Information about social and environmental consequences</i>	Present a speech given by a high status professional to emphasise the importance of not exposing patients to unnecessary radiation by ordering x-rays for back pain
-----	------------------------	---	---

Example of BCT 9.1 from the online training: “A health professional emphasized the importance of physical activity and healthy eating”

### How does this relate to the PA app assessment?

From Lou Atkins (I have asked her for clarification on how to distinguish the presence/absence of this BCT. The below instruction help to clarify the decision on the BCT 9.1 Credible Source).

“It depends how the information is presented in the app. For example, just putting links

to studies is not enough to interpret as credible source. However, if in the app it said something like “the recommendations in this app are based on the scientific literature which advises, e.g. exercising 7 minutes a day is good for you! Click here to find out more about the evidence for exercising for 7 minutes”, then it would be interpreted as credible source.” (Lou Atkins)

Note that the information can be written as well (not only ‘verbal or visual communication’ as in the definition of this BCT).

Further examples of clarification between the presence/absence of 9.1 – Credible Source:

9.1 - Credible Source	App developed by NSCA-Certified Personal Trainer® (NSCA-CPT®)	7 Minute Workout Challenge
NOT Credible Source	Merely giving links to peer-reviewed studies, e.g., Diabetic association and American College of Sports Medicine	Quick Fit - 7

## 5.1 Information about health consequences in app assessment

5.1	<i>Information about health consequences</i>	Provide information (e.g. written, verbal, visual) about health consequences of performing the behavior <i>Note: consequences can be for any target, not just the recipient(s) of the intervention; emphasizing importance of consequences is not sufficient; if information about emotional consequences, code 5.6, Information about emotional consequences; if about social, environmental or unspecified consequences code 5.3, Information about social and environmental consequences</i>	Explain that not finishing a course of antibiotics can increase susceptibility to future infection  Present the likelihood of contracting a sexually transmitted infection following unprotected sexual behavior
-----	--	--	--

It includes both POSITIVE (BENEFITS) AND NEGATIVE (CONSEQUENCES)

Example of BCT 5.1 from the training: “A successful sport personality emphasizes the health benefits of jogging to participants” (Note that 9.1 Credible Source is also present in this example)

We discussed this before and we thought that when health consequences are positive (i.e., benefits) then we will not include these as a BCT. However, it turns out that we need to.

However, emphasizing the important of doing exercise is not enough for the BCT to be present, example from the training:

“The importance of drinking water was emphasized to participants” - “emphasizing the importance” is not enough for the BCT to be present but “emphasizing the health benefits” is.

## 1.1 Goal setting (behaviour)

1.1	<i>Goal setting (behavior)</i>	Set or agree on a goal defined in terms of the behavior to be achieved <i>Note: only code goal-setting if there is sufficient evidence that goal set as part of intervention; if goal unspecified or a behavioral outcome, code 1.3, Goal setting (outcome); if the goal defines a specific context, frequency, duration or intensity for the behavior, <u>also</u> code 1.4, Action planning</i>	Agree on a daily walking goal (e.g. 3 miles) with the person and reach agreement about the goal  Set the goal of eating 5 pieces of fruit per day as specified in public health guidelines
-----	--------------------------------	--	--

## 1.4 Action Planning

1.4	<i>Action planning</i>	Prompt detailed planning of performance of the behavior (must include at least one of context, frequency, duration and intensity). Context may be environmental (physical or social) or internal (physical, emotional or cognitive) (includes <b>Implementation Intentions</b> ) <i>Note: evidence of action planning does not necessarily imply goal setting, only code latter if sufficient evidence</i>	Encourage a plan to carry condoms when going out socially at weekends  Prompt planning the performance of a particular physical activity (e.g. running) at a particular time (e.g. before work) on certain days of the week
-----	------------------------	---	---

We had some inconsistencies in trying to distinguish between the BCTs 1.1 Goal setting (behaviour) and 1.4 Action Planning.

Below is an example of 1.1 Goal setting ONLY:

“Participants agreed to reduce fat content in their diet”. Justification: Participants agreed to a goal defined in terms of the target behaviour”

1.1 Goal setting is less specific than 1.4 Action Planning. For 1.4 Action Planning, there needs to be specified at least one of these:

- context
- frequency
- duration
- intensity

Below is an example of both, 1.1 Goal setting (behaviour) and 1.4 Action Planning.

Example of BCTs 1.1 and 1.4 from the training:

“Participants agreed to take a ten minute walk every time that they felt anxious or stressed.” – Justification: both 1.1 goal setting (behaviour) and 1.4 Action planning – because goal is being set (1.1) and a detailed plan is being agreed “if-then” plan.

A couple more examples of both, 1.1 Goal setting (behaviour) and 1.4 Action Planning:

“The goal was set at 4,000 steps per day, which is equivalent to a half hour walk at a pace of 15 min/mile”. Justification: The goal is being set (1.1) in relation to the target behaviour. Detailed planning is specified which includes how often (frequency) and how long the participant should walk for (duration), and the speed of walking (the intensity). Hence, Action Planning (4.1) also coded.

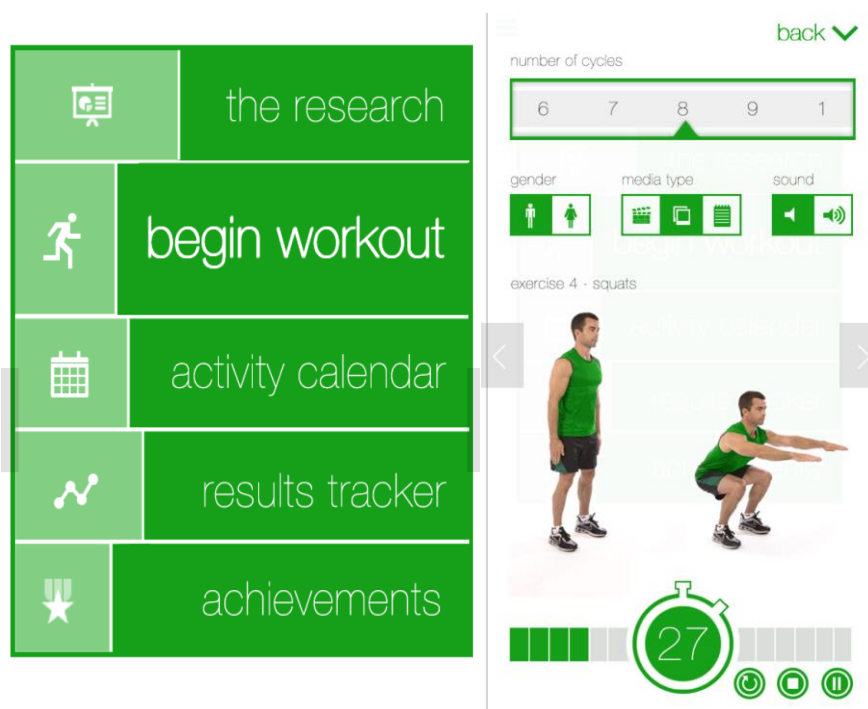
“Medical students agreed to wash their hands correctly after touching each patient.”  
Justification: The students are agreeing to a goal (1.1) in terms of the target behaviour (i.e., hand washing). Action planning is also coded as the goal includes context (i.e., touching the patients) and “if-then plan” – “If I touch the patient, then I will wash my hands”.

## How does this relate to the PA app assessment?

### Example of the presence of 1.1 BCT – Goal Setting (but not 1.4 Action Planning)

App types: 7 minute workouts

Example below from: 7 Minute Workout Challenge



The PA in these apps consists of a 7 minute routine (12 exercises with 10 sec break in-between). The description says to do at least 1 set.

The rationale for the presence of the 1.1 BCT in the above app. It is consistent with the definition from the taxonomy: set or agree on a goal defined in terms of the behaviour to be achieved. The behaviour to be achieved is defined: 12 exercises (agreed with Lou Atkins).

However, this app DOES NOT have 1.4 Action Planning BCT in it.



## Further examples of the presence of both 1.1 BCT – Goal Setting AND 1.4 Action Planning

App types: 0 to 5k

Example from: Coach to 5K



This app sets a goal for training 30 minutes/3 times a week for 9 weeks. By the end of the 9<sup>th</sup> week the user will be able to run for 30 min without stopping.

The BCT: 1.1 Goal setting is present here but also 1.4 Action Planning, i.e., the duration, frequency is set.

The rationale is that the presence of the 1.4 BCT in the Couch to 5K: Apps type, like Couch to 5K plans for the user to do the exercise three times a week. It also schedules the duration of the run and walk at each session until week 9 (the action planning needs to be finite) in which the user is to run 30 min without stopping. I think this is action planning. The app plans it for you and this is really the essence of this app.

## Feedback on behaviour and 2.7 Feedback on outcome(s) of behaviour VERSUS (2.3. Self-monitoring of behaviour OR 2.4. Self-monitoring of outcome(s) of behaviour)

2.2	<b>Feedback on behavior</b>	Monitor and provide informative or evaluative feedback on performance of the behavior (e.g. form, frequency, duration, intensity) <i>Note: if Biofeedback, code only 2.6, Biofeedback and not 2.2, Feedback on behavior; if feedback is on outcome(s) of behavior, code 2.7, Feedback on outcome(s) of behavior; if there is no clear evidence that feedback was given, code 2.1, Monitoring of behavior by others without feedback; if feedback on behaviour is evaluative e.g. praise, also code 10.4, Social reward</i>	Inform the person of how many steps they walked each day (as recorded on a pedometer) or how many calories they ate each day (based on a food consumption questionnaire).
2.7	<b>Feedback on outcome(s) of behavior</b>	Monitor and provide feedback on the outcome of performance of the behavior <i>Note: if Biofeedback, code only 2.6, Biofeedback and not 2.7, Feedback on outcome(s) of behavior; if feedback is on behavior code 2.2, Feedback on behavior; if there is no clear evidence that feedback was given code 2.5, Monitoring outcome(s) of behavior by others without feedback; if feedback on behaviour is evaluative e.g. praise, also code 10.4, Social reward</i>	Inform the person of how much weight they have lost following the implementation of a new exercise regime

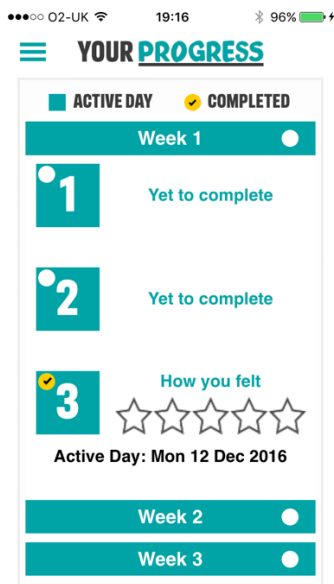
Example of BCT 2.2 from the training: “Students were observed washing their hands and afterwards the instructor discussed their technique”. Justification: The instructor is “monitoring” the students’ hand washing behaviour and is giving them evaluative feedback on their technique.

### How does this relate to the PA app assessment?

When the monitoring is done automatically, e.g., using pedometer or recording exercise routine based on the automatic monitoring system (like in the example of 7 Minute Workout Challenge) then a feedback is given in a form of information (may or may not include a graph) or evaluation then the BCT 2.2 Feedback on behaviour is present.

### What to do when the feedback is very limited?

For example, it is only one very simple feedback, like a tick box that the exercise session was done, then indicate: “+”. Example of this is given below (source: One You Couch to 5k )



Second example of Feedback on behaviour – “+” (when the feedback is limited), GA example:

When the app says (in Challenge-focused apps) - your progress is 30%. It doesn't say for example how many sets of exercise or time spent on it.

If the user must record the information manually then it is BCT 2.3 Self-monitoring of behaviour (PA) or 2.4 Self-monitoring of outcomes of behaviour (weight etc). In this case, a graph itself from the user-inputted information will be a 2.7 Feedback on outcomes of behaviour (for example, a graph of weight loss generated by user-inputted weight over time) but the level of confidence will be “+”. Example below:

7 Minute workout challenge	2.7	Feedback on outcome(s) of behaviour	+	Graph - Weight log (inputted by the user)
----------------------------	-----	-------------------------------------	---	---

Note that when the user inputs weight and height by themselves but the app counts the BMI then the BMI is 2.7 Feedback on outcomes of behaviour.

Timer-type apps are 2.2 Feedback on behaviour (Monitor and provide informative or evaluative feedback on performance of the behaviour (e.g. *form, frequency, duration, intensity*) – as per Lou’s email

2.3	<b>Self-monitoring of behavior</b>	Establish a method for the person to monitor and record their behavior(s) as part of a behavior change strategy <i>Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behavior, do not code; if monitoring of outcome of behavior, code 2.4, Self-monitoring of outcome(s) of behavior; if monitoring is by someone else (without feedback), code 2.1, Monitoring of behavior by others without feedback</i>	Ask the person to record daily, in a diary, whether they have brushed their teeth for at least two minutes before going to bed  Give patient a pedometer and a form for recording daily total number of steps
2.4	<b>Self-monitoring of outcome(s) of behavior</b>	Establish a method for the person to monitor and record the <b>outcome(s)</b> of their behavior as part of a behavior change strategy <i>Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behavior, do not code ; if monitoring behavior, code 2.3, Self-monitoring of behavior; if monitoring is by someone else (without feedback), code 2.5, Monitoring outcome(s) of behavior by others without feedback</i>	Ask the person to weigh themselves at the end of each day, over a two week period, and record their daily weight on a graph to increase exercise behaviors

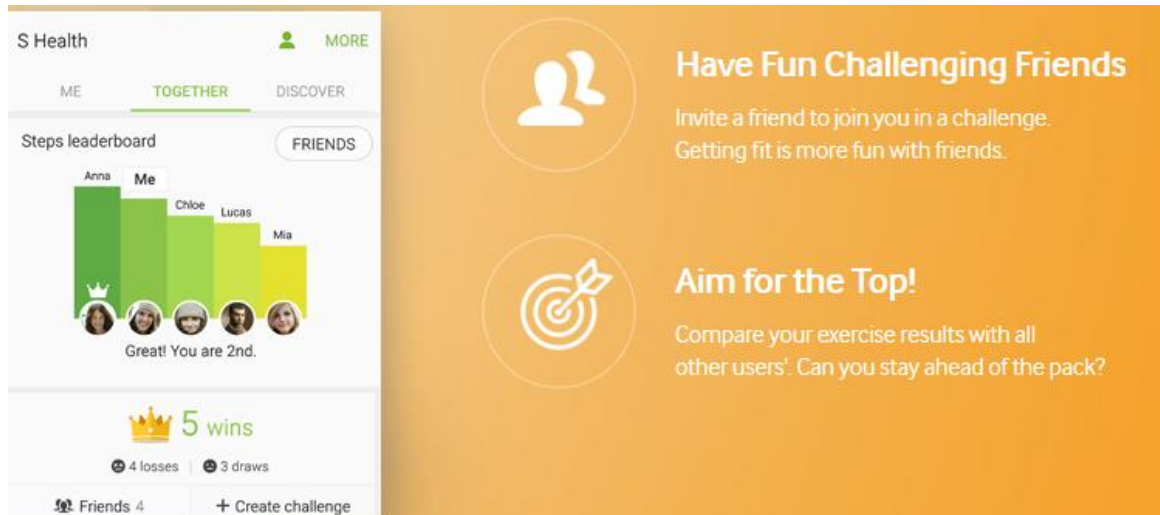
### 3.1 Social Support (unspecified) versus 6.2 Social Comparison

3.1	<b>Social support (unspecified)</b>	Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues, 'buddies' or staff) or non-contingent praise or reward for performance of the behavior. It includes encouragement and counselling, but only when it is directed at the behavior <i>Note: attending a group class and/or mention of 'follow-up' does not necessarily apply this BCT, support must be explicitly mentioned; if practical, code 3.2, Social support (practical); if emotional, code 3.3, Social support (emotional) (includes 'Motivational interviewing' and 'Cognitive Behavioral Therapy')</i>	Advise the person to call a 'buddy' when they experience an urge to smoke  Arrange for a housemate to encourage continuation with the behavior change programme  Give information about a self-help group that offers support for the behaviour
6.2	<b>Social comparison</b>	Draw attention to others' performance to allow comparison with the person's own performance <i>Note: being in a group setting does not necessarily mean that social comparison is actually taking place</i>	Show the doctor the proportion of patients who were prescribed antibiotics for a common cold by other doctors and compare with their own data

## How does this relate to the PA app assessment?

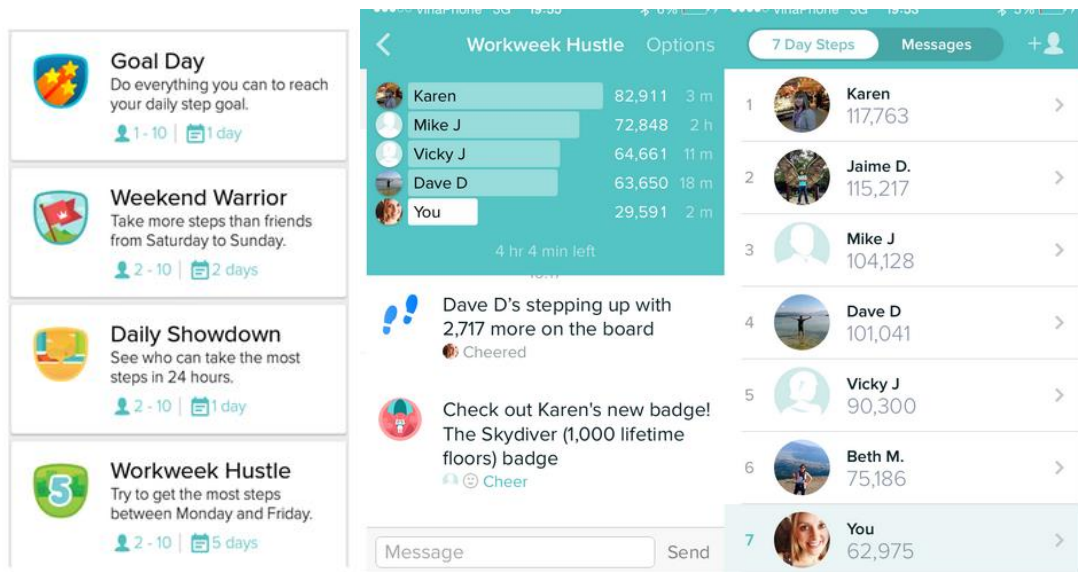
Below are two examples of 6.2 – Social Comparison (source: S Health):

1)



In the second example, the user subscribes to a challenge of up to 10 people and their results are compared (source: Fitbit)

2)



With 6.2 – Social Comparison there needs to be an informed ability to see and compare the PA sessions with your friends' and colleagues or some kind of pre-defined group. Merely posting the users' training results on social media is not enough. There needs to be an element of comparison.

Example of Social Support 3.1 would be a forum for users to share experiences etc. (as per Lou's email)

The table below shows examples of some BCTs from Conroy’s study (attached in the email before, please let me know if I should re-send it).

**Incentives vs rewards, social support (in red is the BCTs that Lou did not agree with). Please adhere to what Lou said.**

Incentives & Rewards (see the text below the table for an example from S Health)	BCT	Examples from Conroy’s study	Notes
	10.1 Material incentive (behaviour)	Informs user that the premium version of the app would be free if user completes all sessions and refers friends.	It’s about informing that something will happen if..., about saving money  Ghadah, you indicated this BCT in the 7 minute workout challenge and I agree with you although before we got rid of it. I put it back in.
	10.2 Material reward (behaviour)	Upgrade to premium version of the app (includes additional features) for free by completing all sessions and referring friends.	
	10.6 Non-specific incentive	Informing a user that more badges could be earned if the user completes a more advanced workout.	It’s about informing that something will happen if the behaviour occurs, e.g., badges, medals, etc. <u>If it is not told in advance then code as 10.3</u>
	10.3 Non-specific reward	Earn badges for each goal achieved.	<u>getting badges, medals, points but only if the user was not informed about provision of a reward. Otherwise code as 10.6</u>
<p>Lou Atkins’ clarification of the difference between 10.6 and 10.3: a reward is something given when there has been effort towards performing the behaviour, it would be coded as incentive if the person is informed that there will be a reward if they make progress towards performing the behaviour.</p> <p>PLEASE MAKE SURE TO RUN AN EXERCISE SO WE CAN FIND OUT IF THERE IS A REWARD GIVEN AFTER COMPLETION OF AN EXERCISE</p>			
	10.4 Social reward	Get verbal encouragement from the audio coach after completing a workout	Needs to be AUDIO, i.e., voice. Example: in 7 minute workout challenge the audio coach says: “workout



			complete. Congratulations” (verified with Lou Atkins who agreed)
	10.5. Social incentive	Get verbal encouragement from the audio coach to motivate them.	As above – it needs to be AUDIO. , i.e., voice. Example: “You can do it!” in One You running app
Social support	3.1 Social support (unspecified)	An option to share your progress and personal records with friends via Facebook, Twitter, or email; or within the app community	Note: The app needs to share the PA session recorded by that app. If it just says something like: I have just completed a workout using this app” then we treat this as an advertisement, not as BCT 3.1  Posts like: "I've just finished run for 48 min" “this is day 1 of the 100 day challenge I am doing”  are examples of Social Support (unspecified) ++ Rationale in notebook 6/1/17
	3.2 Social support (practical)	Provide online coach to help the user complete each training.  NOT AGREED (Lou Atkins) – this is Instruction on how to perform behaviour	

### 3.1 Social support (unspecified) – the case of user-generated forums

2. FORUM - For a user-generated forum where any topics can be discussed - indicate 3.1 Social support (unspecified)

I came across a forum – very basic with user-generated content. The content was various: different advice, some emotional support, practical support and other. I was really unsure what to put in there, especially that it seemed that the developers had little over what is in the forum. I decided to ascribe the forum BCT (the most general BCT) - Social support (unspecified). This is because of the Principle 3 (can be found below) of the online BCT training which you have done as well, you mentioned. This is from the Manual. Please, could you read it to remind yourself about it.

### 3.2 Social support (practical)

If the app helps to sign up for REAL LIFE events – then it is social support (practical)

### 6.3 Information about others’ approval

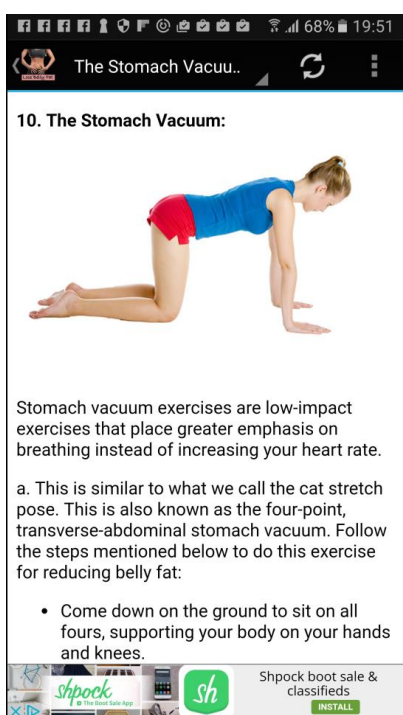
Liking the entry about the exercise session on the friend's feed (heart, like, clap). ++

## 7.1 Prompts/Cues

Notifications does not mean that they are reminders – they are reminders if they remind about the behaviour to be completed. Please list the feature as NOTIFICATIONS and write what the notifications do if known. If they are reminders as well than write Reminders – in the sub-feature (plus the BCT).

## Demonstration of the behaviour'

One photo does not constitute - 6.1 Demonstration of behaviour. It needs to be at least changing image to demonstrate how to perform a behaviour. If there is a photo and text then it is 4.1 Instruction of how to perform a behaviour. Example of 4.1 and NOT 6.1 below (source: Belly Fat Exercises):



## Difference between an INCENTIVE and a REWARD

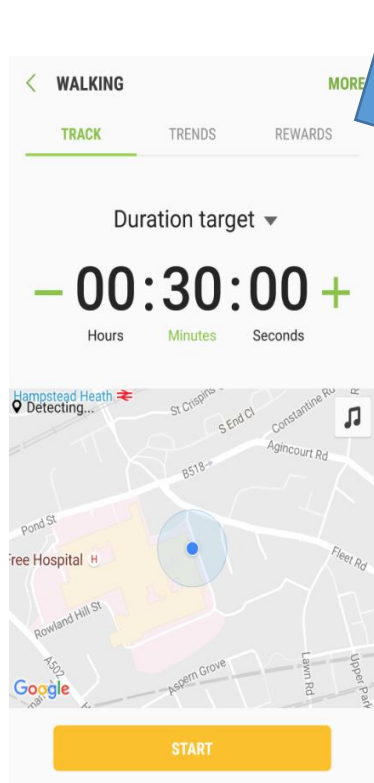
If there is NO sections on rewards, achievements, medals, etc., and then the reward was provided then code it as a REWARD

If there is a sections on rewards, achievements, medals, etc., then code it as an INCENTIVE. Please make sure that you explore the whole app to see if there is a section on rewards, achievements, medals, etc.,

Please remember: if money is involved then consider MATERIAL INCENTIVE (see the table above



Below is an example of an INCENTIVE as there is a section that informs you about provision of rewards.



This is an incentive as it is known that reward will be provided if the behaviour is completed.

### Some notes on the Principle 3

1. Take care distinguishing between BCTs that only differ in terms of their behaviour change type (i.e. behaviour vs. outcome).

'Goal setting (behaviour)' 1.3 'Goal setting (outcome)'

Example from the training: "To reduce participant's alcohol consumption, goal setting was prompted"

The BCT type (behaviour vs. outcome) is unspecified in the example above. In this context, coders are instructed to code this BCT as 1.3 Goal setting (outcome) type. This is because if one were evaluating BCT effectiveness the outcome version is less likely to over-estimate BCT effectiveness.

The same rule applies to the following 6 groupings of BCTs in the training taxonomy that include behaviour and outcome types:

- BCTs: 'Feedback on behaviour' and 'Feedback on outcome(s) of behaviour'
- BCTs: 'Self-monitoring of behaviour' and 'Self-monitoring of outcome(s) of behaviour'
- BCTs: 'Monitoring of behaviour by others without feedback' and 'Monitoring outcome(s) of behaviour by others without feedback'

- BCTs: 'Material reward' and 'Material reward (outcome)'
- BCTs: 'Goal setting (behaviour)' and 'Goal setting (outcome)'
- BCTs: 'Review behaviour goal(s)' and 'Review outcome goal(s)'

The following three groupings of BCTs differ in type for other reasons:

- BCTs: 'Social support (practical)' and 'Social support (unspecified)'
- BCTs: 'Material reward', 'Social reward' and 'Non-specific reward'
- BCTs: 'Information about health consequences', 'Information about emotional consequences' and 'Information about social and environmental consequences' (Notice: the taxonomy doesn't distinguish between outcomes and behaviours so can be used or both)

Example from the training: “Participants were provided information on the consequences of giving up meat”.

In this phrase the BCT type is unspecified and cannot reasonably be inferred (see: Learning principle 2). In this context coders are instructed to code the most general BCT type in the particular grouping as present. In this illustration the BCT: ‘Information about social and environmental consequences’ should be coded as present.

Imagine later in the description, the following phrase: “Participants were told that giving up meat may lead to a deficiency in vitamin B12”

This phrase clarifies that the consequences refer to health consequences and therefore only the BCT ‘Information about health consequences’ should be coded as present.

#### Notes on the Principle 4

1. Code technical terms and packages of BCTs that map onto BCTs in the taxonomy

Technical terms and BCT packages that are reasonably well defined (i.e. they consistently include the same collection of BCTs) are included underneath BCT definitions in the taxonomy.

#### Summary of technical terms / BCT packages:

- 'Extinction' - code as the BCT 'Remove reward'
- 'Relapse prevention' - code as the BCT 'Problem solving'
- 'Coping planning' - code as the BCT 'Problem solving'
- 'Implementation intentions' - code as the BCT 'Action planning'
- 'Cognitive restructuring' - code as the BCT 'Framing / reframing'
- 'Modelling' - code as the BCT 'Demonstration of the behaviour'
- 'Positive reinforcement' - code as the BCT/s 'Material reward (behaviour)' and/or 'Material reward (outcome)' and/or 'Social reward' and/or 'Non-specific reward'
- 'Skills training' - code as the BCT 'Instruction on how to perform the behaviour'
- 'Stress management' - code as the BCT 'Reduce negative emotions'
- 'Decisional balance' - code as the BCT 'Pros and cons'

Appendix F: Individual-level data for the sample of apps assessed (Study 1)

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Fitbit	TM	7871 / 207000	FREE	104 / 22.35	09/12/2016 - 15/12/2016	●	●	●	●	○	●	●	10	○	●	3.9 / 4
Strava Running and Cycling GPS	TM	10742 / 199596	FREE	59.2 / 21.41	09/01/2017 - 03/01/2017	●	●	●	●	○	●	●	8	○	●	4.6 / 4.5
Pacer – Pedometer plus Weight Loss and BMI Tracker	P	4484 / 250669	FREE	85.4 / 9.64	12/01/2017 - 09/01/2017	●	●	●	●	○	●	●	10	○	●	4.6 / 4.5
Map My Run – GPS Running & Workout Tracker	TM	24530 / 135087	FREE	164 / 70.02	22/12/2016 - 22/12/2016	●	●	●	○	○	●	●	8	○	●	4.5 / 4.5

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Adidas train & run	TM	NA / 78868	FREE	NA / 72.53	NA - 30/11/2016	●	●	●	●	●	●	●	10	○	○	NA / 4.3
Steps Pedometer & Step Counter Activity Tracker	P	438 / 0	FREE	25.7 / NA	30/11/2015 - 07/12/2016	○	○	○	○	○	○	○	3	○	○	4.1 / NA
7 Minute Workout by Simple Design Ltd	W	NA / 351356	FREE	NA / 13.67	19/12/2016 - 19/12/2016	●	●	●	NA	○	○	○	11	○	○	NA / 4.5
Runtastic Running & Fitness	TM	1760 / 625077	FREE	144 / 33.08	NA - 13/12/2016	●	●	○	○	○	○	●	8	○	●	4.5 / 4.5

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Home workout MMA Spartan Free	W	NA / 9913	FREE	NA / 28.41	12/01/2017 - NA	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	5	<input type="radio"/>	<input type="radio"/>	NA / 4.6
Stepz – Pedometer & Step Counter	P	387 / 0	FREE	43.4 / NA	27/05/2016 - NA	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	8	<input type="radio"/>	<input type="radio"/>	4.6 / NA
Interval Timer – Timing for HIIT Training and Workouts	IT	119 / 0	FREE	14.3 / NA	17/01/2017 - NA	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	2	<input type="radio"/>	<input type="radio"/>	4.4 / NA

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Sworkit – Custom Workouts for Exercise & Fitness	W	1383 / 0	FREE	151 / NA	NA - 20/10/ 2016	●	●	●	○	○	●	●	8	●	●	4.7 / NA
Fitness & Bodybuilding	W	NA / 44923	FREE	NA / 31.5	29/01/2016 - NA	●	●	●	○	○	●	●	7	●	○	NA / 4.6
Daily Workouts FREE	W	826 / 0	FREE	52 / NA	03/11/2016 - 21/06/2016	●	●	●	○	○	●	●	7	●	○	4.6 / NA
30 Day Ab Challenge FREE	W	498 / 996	FREE	67.7 / 6.72	12/01/2017 - NA	○	○	○	○	○	○	○	5	○	○	3.7 / 4.5

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Runtastic Results: Body Workout Fitness Trainer	W	1763 / 0	FREE	163 / NA	NA - 12/12/2016	●	●	○	○	○	●	●	5	○	○	4.5 / NA
C25K® - 5K Running Trainer	RP	NA / 29795	FREE	NA / 38.53	03/01/2017 - NA	●	●	○	○	○	●	●	7	○	●	NA / 4.6
Health Mate – Steps tracker & Life coach	P	758 / NA	FREE	124 / NA	03/08/2016 - NA	●	●	●	○	●	●	●	8	○	○	3.2 / NA
One You Couch to 5K	RP	92 / 206	FREE	73.9 / 33.07	01/08/2016 - NA	●	●	●	○	○	●	●	12	○	○	1.8 / 2.4

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Running, Walking and Biking with Endomondo	TM	3698 / 377971	FREE	131 / 18.24	15/01/2017 - 15/12/2016	●	●	○	○	○	●	●	7	○	●	4.4 / 4.5
Map My Ride – GPS Cycling & Route Tracker	TM	10197 / 78204	FREE	161 / 68.09	21/12/2016 - 22/12/2016	●	●	●	○	○	●	●	8	○	●	4.2 / 4.4
Interval Timer	IT	NA / 5827	FREE	NA / 4.82	NA - 24/04/2016	○	○	○	○	○	○	○	1	○	○	NA / 4.7
5K Run - Couch to 5K	RP	NA / 1615	FREE	NA / 26.82	NA - 29/12/2016	●	●	○	○	○	●	●	6	○	○	NA / 4.3



Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
7 Minutes Workout – Women Fitness Exercise Trainer	W	110 / NA	FREE	92.7 / NA	18/09/2016 - NA	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	8	<input type="radio"/>	<input type="radio"/>	4.3 / NA
Seconds – Interval Timer	IT	633 / NA	FREE	11.2 / NA	23/11/2015 - NA	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	1	<input type="radio"/>	<input type="radio"/>	4.6 / NA
Running Distance Tracker +	TM	NA / 4116	FREE	NA / 12.15	NA -11/01/2017	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	NA	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	3	<input type="radio"/>	<input type="radio"/>	NA / 4.6
Freeletics Bodyweight – Workout	W	602 / 77767	FREE	70.6 / 38.97	12/01/2017 - 11/01/2017	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	11	<input type="radio"/>	<input type="radio"/>	4 / 4.5

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Couch to 10K Running Trainer	RP	NA / 5856	FREE	NA/ 40.06	NA -16/01/2017	●	●	○	○	○	●	●	7	○	○	NA / 4.6
FitNotes - Gym Workout Log	W	NA / 12698	FREE	NA/ 1.88MB	NA -30/11/2016	○	○	○	○	○	○	○	6	○	○	NA / 4.5
Belly Fat Exercises	W	NA / 1475	FREE	NA/ 3.93MB	NA -30/07/2016	○	○	○	○	○	○	○	1	○	○	NA / 3.9
Belly Fat Workout FREE – 10 Minute Ab Exercises	W	672 / NA	FREE	74.3 /NA	18/11/2016 - NA	●	●	○	NA	○	NA	NA	5	●	○	4.3 / NA

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Movesum – Step counter by Lifesum	P	14 / NA	FREE	43.6 /NA	24/09/2016 - NA	●	●	○	○	○	●	□	4	○	○	1.9 / NA
7 Minute Workout Challenge by Fitness Guide Inc	W	1247/3697	2.29/2.48	102 / 43.88	20/04/2015 - 23/02/2015	○	○	○	○	○	○	○	11	●	○	4.7 / 4.4
Adrian James 6 Pack Abs Workout	W	801/NA	0.79 / NA	73.5 /NA	20/09/2016 - NA	●	●	○	○	○	●	○	9	●	○	4.6 / NA
Full Fitness : Exercise Workout Trainer	W	1455 / NA	2.29 / NA	105 /NA	20/12/2016 - NA	○	○	○	○	○	○	○	9	○	○	4.4 / NA

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Runtastic PRO Running and Workout Tracker	TM	3845 / 177277	3.99 / 4.99	143 / 33.3	17/01/2017 - 22/12/2016	●	●	○	○	○	●	●	9	○	○	4.6 / 4.5
Couch to 5K® - Running App and Training Coach	RP	127 / 9207	1.49 / 1.67	97.2 / 15.4	30/11/2016 - 18/11/2016	●	●	●	○	○	●	●	9	○	○	4.3 / 4.6
Adrian James High Intensity Interval Training	W	77 / NA	0.79 / NA	82.2 / NA	20/09/2016 - NA	●	●	○	○	○	●	○	7	●	○	4.8 / NA

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Running for Weight Loss PRO	TM	3825 / NA	1.49 / NA	104 /NA	16/12/2016 - NA	●	●	○	○	○	●	●	10	○	○	4.6 / NA
Fitness Trainer FULL version	W	NA / 2876	NA / 0.99	NA/ 52.36	NA -29/12/2016	○	○	○	○	○	○	○	7	○	○	NA / 4.7
Instant Fitness : 600+ exercises, 100+ workouts...	W	79 / NA	2.29 / NA	80.2 /NA	30/05/2014 - NA	●	●	○	○	○	●	●	7	○	○	4.3 / NA
Push ups 0 to 100: push ups	W	106 / NA	1.49 / NA	65.6 /NA	06/10/2016 - NA	●	●	○	NA	○	○	○	9	○	○	4.8 / NA

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
challenge trainer pro																
Couch to 5K Runner, 0 to 5K run training	RP	1678 / NA	2.29 / NA	84.3 / NA	04/10/2016 - NA	●	●	○	NA	○	○	○	10	○	○	4.9 / NA
Footsteps – Pedometer	P	221 / NA	2.29 / NA	9.3 / NA	13/12/2015 - NA	●	●	○	NA	○	○	○	3	○	○	3.4 / NA
iMuscle 2	W	56 / 2128	2.79 / 1.89	376 / 145	08/12/2014 - 22/07/2016	●	●	○	○	○	●	●	10	●	○	4.8 / 4.1
10K Running Trainer Pro	RP	NA / 570	NA / 2.49	NA / 27.64	NA -09/01/2017	●	●	○	○	○	●	●	8	○	○	NA / 4.5

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Police Fitness – Bleep Test	W	20 / NA	0.79 / NA	30.6 / NA	27/01/2017 - NA	○	○	○	○	○	○	○	3	○	○	4 / NA
Marathon Trainer - 26.2 42K	RP	NA / 136	NA / 8.49	NA / 27.66	NA -10/01/2017	●	●	○	○	○	●	●	8	○	○	NA / 4.4
MapMyFitness+ Workout Trainer	W	NA / 1409	NA / 2.30	NA / 88.23	NA - 31/01/2017	●	●	●	○	●	●	●	10	○	●	NA / 4.4
5K to 10K	RP	16 / 378	NA 2.29 / 2.32	79.4 / 12.92	30/12/2015 - 25/05/2016	●	●	●	○	○	●	●	10	○	○	3.9 / 4.3
7 Minute Workout Pro	W	NA / 1521	NA / 2.39	NA / 28.2	NA -26/09/2016	●	●	●	NA	○	○	○	13	○	○	NA / 4.6

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Chloe Madeley Weights 4 Women	W	11 / 7	NA2.99 / 2.99	102 / 60.19	03/05/2016 - 04/05/2016	●	●	●	○	○	●	●	2	●	○	2.5 / 3.9
10K Pacer: Run pace training. Run faster	RP	12 / NA	NA2.99 / NA	65.3 / NA	08/10/2016 - NA	●	●	○	NA	○	○	○	10	○	○	4.8 / NA
Starting Strength Official	W	NA / 391	NA / 8.99	NA / 55.58	NA -15/12/2016	○	○	○	○	○	○	○	9	●	○	NA / 4.6
Thor Fitness: 60 Day Bodyweight Workout Routine	W	17 / NA	0.79 / NA	242 / NA	10/06/2015 - NA	○	○	○	○	○	○	○	8	○	○	4.4 / N



Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Half Marathon Trainer 13.1 21K	RP	47 / NA	7.99 / NA	90.7 / NA	06/01/2017 - NA	●	●	○	○	○	●	●	8	○	●	4.3 / NA
Yoga Break	W	not displayed / NA	2.29 / NA	98.5 / NA	09/02/2016 - NA	○	□	□	□	□	□	□	5	○	○	not displayed / NA
PDC Pole Dance Syllabus	W	NA / 15	NA / 2.99	NA / 0.97	NA -14/08/2016	●	●	○	NA	○	●	○	1	●	○	NA / 3.5
MMA Spartan Workouts Pro	W	NA / 1920	NA / 2.99	NA / 44.07	NA -21/09/2016	○	○	○	○	○	○	○	6	○	○	NA / 4.6
Get Running	RP	2670 / NA	2.29 / NA	24.2 / NA	16/10/2013 - NA	○	○	○	○	○	○	○	5	○	●	4.7 / NA

Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
(Coach to 5K)																
WalkJogRun GPS Running Routes	TM	935 / NA	3.99 / NA	44.8 / NA	04/10/2016 - NA	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	5	<input type="radio"/>	<input type="radio"/>	3.9 / NA
Runtastic Road Bike PRO	TM	NA / 14840	NA / 4.99	104 / 22.35	NA -01/03/2016	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	7	<input type="radio"/>	<input type="radio"/>	NA / 4.4
Runtastic Mountain Bike PRO GPS Biking Computer, Trail and Route Tracker	TM	232 / 14955	3.99 / 4.99	59.2 / 21.41	12/09/2016 - 01/03/2016	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	6	<input type="radio"/>	<input type="radio"/>	4.8 / 4.5

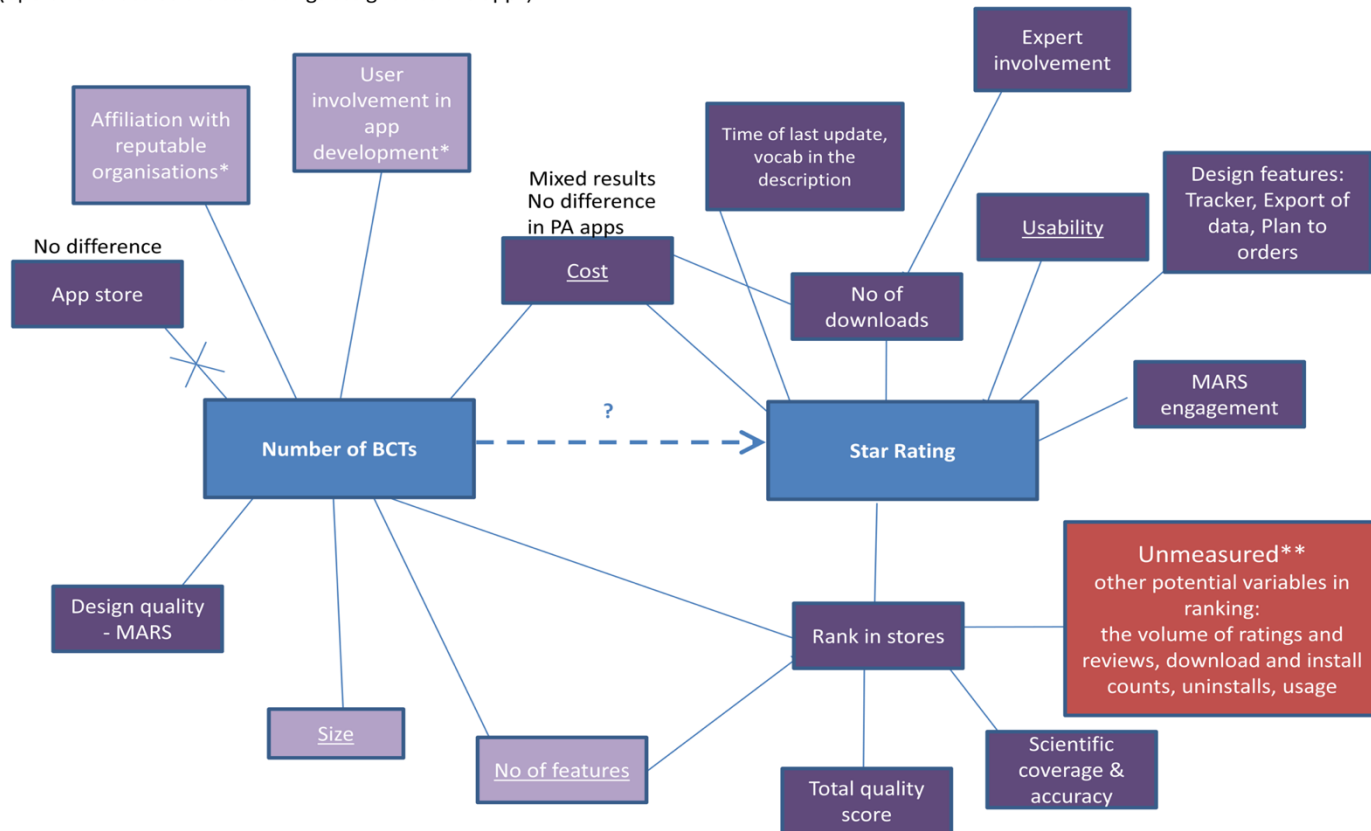
Name	Type of PA app	Number of ratings (iTunes / GP)	Cost (iTunes / GP)	Size in MB (iTunes / GP)	Last update (iTunes / GP)	Have privacy policy	Available without app download	Available after app download	Short form notice	Other languages	Collects PII	Shares with 3rd party	Total number of BCTs	Expert involvement	Any studies associated with the app	Avg. user rating (1-5 stars) (iTunes / GP)
Chloe Madeley 15 minute fat loss workout	W	56 / NA	2.99 / NA	85.4 / 9.64	29/07/2016 - NA	○	○	○	○	○	○	○	3	●	○	3.4 / NA
CARROT Fit – 7 Minute Workout, Step Counter Weight Tracker	W	263 / NA	2.49 / NA	164 / 70.02	17/11/2016 - NA	●	●	●	○	○	○	NS	11	○	○	4.7 / NA

Notes: The assessment of User involvement and Organizational affiliation is not displayed in this table as no apps reported consulting users in the app development; 1 app, One You Couch to 5K, was indicated as having governmental affiliation; App types: IT; Interval timer-type app; P; Pedometer-type app; RP, Running programme-type app; TM; Tracking of movement-type app; W, Workout-type app; MB, megabytes; NA, not applicable; NS, not specified; PPI, Personally Identifiable Information; ●, Yes; ○, No.

Appendix G: Directed Acyclic Graph (Study 2)

a) visualisation of all variables considered

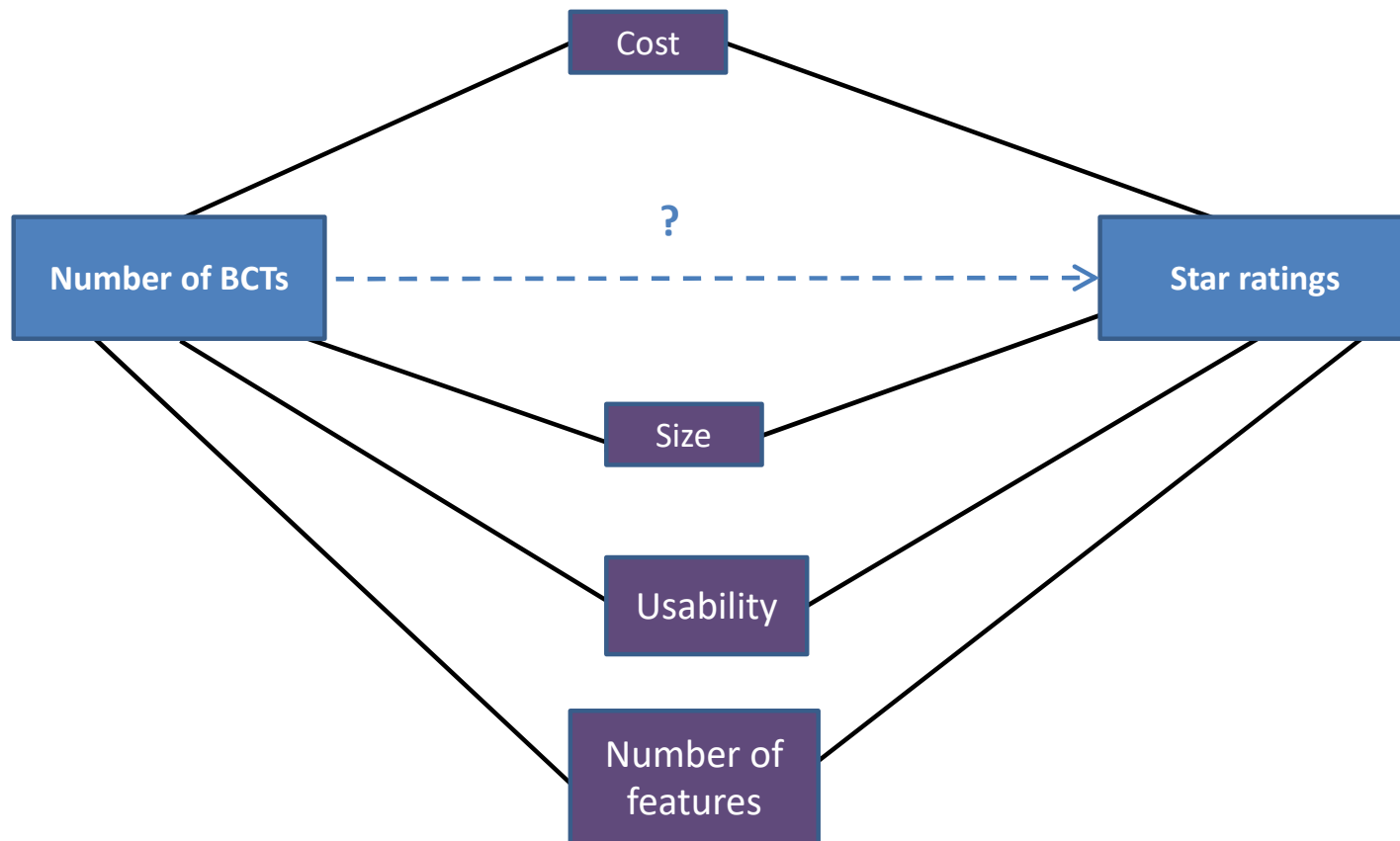
Question of interest: Is there a relationship between the likely efficacy of PA apps (operationalised as the presence of BCTs) and app popularity (operationalised as the star ratings assigned to the apps)?



Notes: Dark colour denotes variables that were assessed in previous studies; \*Light purple colour signifies the association that were not assessed but I considered as potential variables influencing the BCTs and/or user rating; \*\*Red colour describes the variables that we were unable to control for which comprise the ranking algorithm.

a) graph depicting potential confounders that were measured and control for in the analysis

Question of interest: Is there a relationship between the likely efficacy of PA apps (operationalised as the presence of BCTs ) and app popularity (operationalised as the star ratings assigned to the apps)?



Appendix H: Sample size calculation for the association between the number of BCTs and user ratings (Study 2)

A sample size of 51 apps will provide 90% power to detect a change of 0.11 for each additional BCT at 5% significance level (type I error rate).

Input			Results	
$\alpha$	<input type="text" value=".05"/>	Calculate	$\sigma$	<input type="text" value="0.7003"/>
$\beta$	<input type="text" value=".1"/>	Reset	$\rho$	<input type="text" value="0.4157"/>
$\sigma_x$	<input type="text" value="2.91"/>		$N$	<input type="text" value="51"/>
$\sigma_y$	<input type="text" value="0.77"/>			
$\lambda_1$	<input type="text" value="0.11"/>			

A sample size of 65 apps was selected to account for any randomly selected apps that might not fulfil the inclusion criteria once downloaded. If all 65 apps are eligible, the resulting study will have 94% power to detect a change a mean change of 0.11. These calculations assume that pilot data are similar to the study data in that the mean change is 0.11.

Input			Results	
$\alpha$	<input type="text" value=".05"/>	Calculate	$\sigma$	<input type="text" value="0.7272"/>
$\beta$	<input type="text" value=".2"/>	Reset	$\rho$	<input type="text" value="0.328E"/>
$\sigma_x$	<input type="text" value="2.91"/>		$N$	<input type="text" value="65"/>
$\sigma_y$	<input type="text" value="0.77"/>			
$\lambda_1$	<input type="text" value="0.0870"/>			

Reference:

Dupont, W. D., & Plummer Jr, W. D. (1998). Power and sample size calculations for studies involving linear regression. *Controlled clinical trials*, 19(6), 589-601.

Appendix I: Table of different designs explored to assess PA app (Study 3)

<b>Design</b>	<b>Description</b>	<b>Issues raised at the last supervision relating to each design</b>	<b>Notes on the design</b>	<b>Pros</b>
Single Case Design: N-of-1	A type of study from a family of Single Case Designs. A typical single patient trial consists of experimental/control treatment periods repeated a number of times. A typical single patient trial consists of experimental/control treatment periods repeated a number of times. When applied to behaviour change intervention using apps, the use of app (e.g. pedometer) and control period (e.g. using healthy eating app) is randomised each day.	N-of-1 study: EM suggested that this is not an appropriate study because the research method is more used in a drug trial, there needs to be a washout period and it will not be appropriate for a study assessing behavioural intervention.	Random allocation to the trial and control (especially each day) would not be appropriate as: 1) Withdrawal – it would not be beneficial for the users to withdraw them from the PA intervention / also it might not be technically possible as I cannot delete the app from their phone. 2) washout period in behaviour change is an issue	It has a control so I can investigate if the change in PA might be due to the intervention delivered using the apps
Single Case Design with multiple baseline	Baseline assessment is treated like a control. Measurement of the dependent variable during a baseline should occur until the observed	Not discussed	define what is stable PA would be difficult; issues with remotely accessing data from accelerometer	Small sample size needed; more rigorous that pre-post.

	pattern of responding is sufficiently consistent to allow prediction of future responding.			
RCT		PB considered using an RCT but it would be hard to justify assessing only one app from the review (as this will be suggested at her upgrade viva). An RCT with 4 arms plus a control arm would need a large sample size.	There is no equipoise whether PA is beneficial therefore randomising a participant who is interested in increasing their PA (my sample will be people who are motivated but inactive) to a control with no exercise is ethically questionable. At the same time, there is no evidence that the apps I plan to assess actually increase PA. In addition, in this case, active control <i>would be appropriate</i> .	Most rigorous
Factorial design	Factorial design enables to assess various components of the intervention.	Factorial design: my not be appropriate unless PB wanted to assess the use of combination of apps, which is not the research question in this study.	If I was assessing various features of the same app then factorial design would be appropriate but I am assessing 2 apps and 2 behaviours	N/A as not possible for this study.
Pre-post study	A pre-post study examines whether participants in an	A at the meeting occurred that the most appropriate design will be simple pre and post	No control; without reference to a comparison group, it cannot answer whether participants' improvement or	Low-cost, simple, can conduct with a wider range of apps



	<p>intervention improve during the course of the intervention, and then attributes any such improvement or deterioration to the intervention.</p>	<p>single arm design. However PB wants to assess 4 different apps. The design would be to give participants a choice of 4 apps and to conduct a single arm pre and post study. This would be a better option in order to try and mirror choices on the market. People are more likely to use the app/intervention if they are given a choice.</p>	<p>deterioration would have occurred anyway, even without the intervention. It may lead to erroneous conclusions about the effectiveness of the intervention.</p> <p>Issue with interpretation of the results: If the results will show that there is a difference then I can interpret the results as: apps seem to increase, overall, PA.</p> <p>The problem is if there will be no difference from baseline. What would the result mean if users will have a choice of using 1 of 4 apps? That apps for PA do not work? I think it will be difficult to interpret the data</p>	
Latin square 4 x 4	<p>Differs from crossover design in terms of the number of studied treatments; is used when more than two treatments are compared in the same trial.</p>	<p>the issue with “forcing” users to use them all is an issue why Latin square might not work. If some people do not like certain apps then they will not use it and the results might show that they do not work but in fact it might be</p>	<p>It might be complex and a burden on the participants to use 4 apps therefore not the best use of resources.</p>	<p>Would be an opportunity to assess 4 types of apps that are most common in the PA app market</p>

		that they just don't like the particular physical activity provided in the intervention delivered by an app. in addition, the washout period in a behavioural intervention will be an issue.		
Crossover design	Each participant in a crossover trial receives two treatments in a random order and acts as their own control	Still only 2 apps, Unknown how long the baseline assessment of PA should be; I don't have a control group, I am comparing participants to themselves. It might be that I will have a very motivated sample of people only. I'm using the baseline of these people to represent people who I am not studying.	<p>This is the way to simplify 4x4 Latin square. The population group I target in the intervention are people who are motivated to do PA but they do not engage in much PA. Hence they will need to be willing to try the 2 apps I am assessing (easy: walking; slightly harder: 7 minutes workout).</p> <p>This is the most appropriate design for the study to answers some questions with minimal recourses possible (and reduced participant burden)</p>	Ability to assess 2 apps; can assess the overall effect but also the effect of each of the 2 app; Ability to assess the effects of the sequence of app assessment;

Appendix J: Summary of the process of selection of correlates/determinants of PA to measure (Study 3)

<b>Determinants/correlates of PA from Bauman et al. 2012</b>	<b>Mapped onto TDF?</b>	<b>Modifiable/ modifiable in 2-week intervention period?</b>	<b>Measured in the study?</b>
<b><i>Biological/ Demographic</i></b>			
Male sex	Not relevant	non-modifiable	At baseline
White ethnicity	Not relevant	non-modifiable	At baseline
Age (inverse)	Not relevant	non-modifiable	At baseline
Education level	Not relevant	non-modifiable	At baseline
Overweight or obesity	Not relevant	Modifiable but unlikely in the short term	No
<b>Health status</b>	Not relevant	Modifiable but unlikely in the short term	No
<b><i>Psychosocial</i></b>			
<b>Self-efficacy</b>	Measured and mapped	Modifiable	At baseline & follow up
<b>Intention to exercise</b>	Mapped	Modifiable	At baseline & follow up
Stages of readiness to change	Measured and mapped	Modifiable	No, targeting population already motivated to change but doing little or no PA
Perceived effort of the PA	Mapped	Modifiable but unlikely in the short term	No, explored in the qualitative study
Social support	Mapped	Potentially modifiable by the social features of the app	No, explored in the qualitative study
<b><i>Behavioural</i></b>			
<b>History of PA during adulthood</b>	Unable to change/measured, also eligibility criteria	non-modifiable	No, explored in the qualitative study
High job strain	Mapped	non-modifiable	No, explored in the qualitative study alongside other external factors
Action planning	Mapped	Modifiable. The features of app enable self-regulation.	No, explored in the qualitative study

<b>Environmental</b>			
Proximity to facilities	Mapped	non-modifiable	No, explored in the qualitative study
Transportation environment (e.g. pavement and safety of crossings)	Mapped	non-modifiable	No, explored in the qualitative study
Aesthetics	Mapped	non-modifiable	No, explored in the qualitative study
Social environment (safety from crime and social incivilities)	Mapped	non-modifiable	No, explored in the qualitative study
Neighbourhood design (e.g. walkability, street connectivity)	Mapped	non-modifiable	No, explored in the qualitative study

Notes: Bolded items = most consistent correlates, Green = influences that were mapped onto COM-B and TDF, Dark green = measure in the quantitative study

Source of the determinants/correlates:

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., Martin, B. W., & Lancet Physical Activity Series Working Group. (2012). Correlates of physical activity: why are some people physically active and others not?. *The Lancet*, 380(9838), 258-27

Appendix K: Mapping of the determinants/correlates of physical activity (Study 3)

<b>COM-B component</b>		<b>TDF Domain</b>	<b>Relevant theoretical constructs represented within each domain</b>	<b>Determinants/correlates of PA from Bauman et al. 2018</b>
<b>Capability</b>	Psychological	Knowledge	-	
		Skills (cognitive and interpersonal)	-	
		Memory, Attention and Decision Processes	-	
		Behavioural Regulation	Self-monitoring, breaking habits, action planning- provided by app	Action Planning
	Physical	Skills (physical)		Perceived effort of PA
<b>Opportunity</b>	Social environment	Social Influences	Social pressure, social comparisons, group norm, social support, group identity	Social support, Cultural & social norms/ expectations
	Physical environment	Environmental Context and Resources	Material resources - provided by app	Proximity to facilities Transportation environment (e.g. pavement and safety of crossings) Aesthetics Social environment (safety from crime and social incivilities) Neighbourhood design (e.g. walkability, street connectivity) High job strain
<b>Motivation</b>	Reflective	Social/Professional Role & Identity	Identity, group identity	
		Beliefs about Capabilities	Self-confidence, perceived competence, self-efficacy, PBC, beliefs, self-esteem, empowerment	Self-efficacy, Perceived effort of the PA

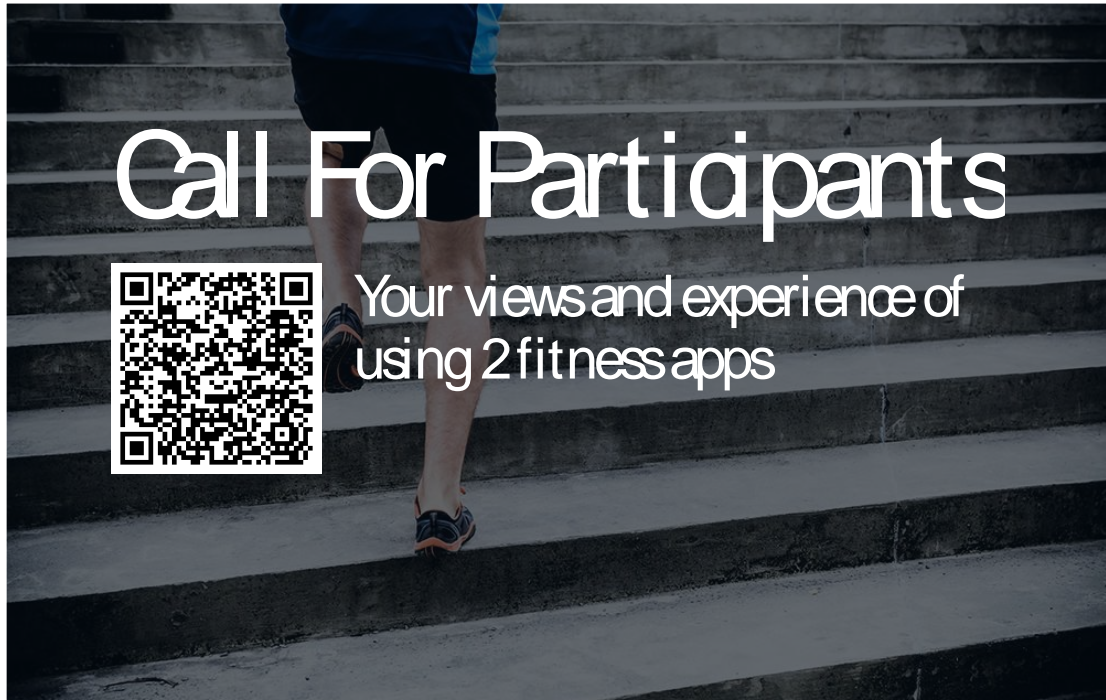
		Optimism	Optimism, identity	
		Beliefs about Consequences	Beliefs, outcome expectancies, characteristics of outcome expectancies, consequents	
		Intentions	Stability of intentions, stages of change model, Transtheoretical model and stages of change	Intention to exercise, Stages of readiness to change
		Goals	Goals (distal/proximal), goal priority, goal/target setting, goals (autonomous/controlled), action planning, implementation intention	Action Planning
	Automatic	Social/Professional Role & Identity	Identity, group identity	
		Optimism	Optimism, identity	
		Reinforcement	Rewards (proximal/distal, valued/not valued, probable/improbable), incentives, consequents, reinforcement, contingencies - provided by app	
		Emotion	Positive/negative affect	

*Reference:*

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., Martin, B. W., & Lancet Physical Activity Series Working Group. (2012). Correlates of physical activity: why are some people physically active and others not?. *The Lancet*, 380(9838), 258-271

Appendix L: Recruitment materials (Study 3)

a) paper advertisement



5 week(s) to complete



£20 gift card, free access to paid apps



Your views and experience of using 2 fitness apps



London, UK

University College London

ARE YOU LOOKING FOR MOTIVATION TO INCREASE YOUR FITNESS?

Would you like a free access to paid mobile apps and a £20 voucher?

Would you consider yourself to be someone who does NOT do enough physical activity?

There are many fitness apps on the market BUT we don't know if any of these apps are helpful in increasing fitness level. Please help us to assess them and increase your physical activity as well!

Find out more online

Poster printed on 28/01/2018 Study expires on 28/02/2018

More info

by scanning the QR code or visiting the URL

[www.dfp.ac/EB6PQ3](http://www.dfp.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

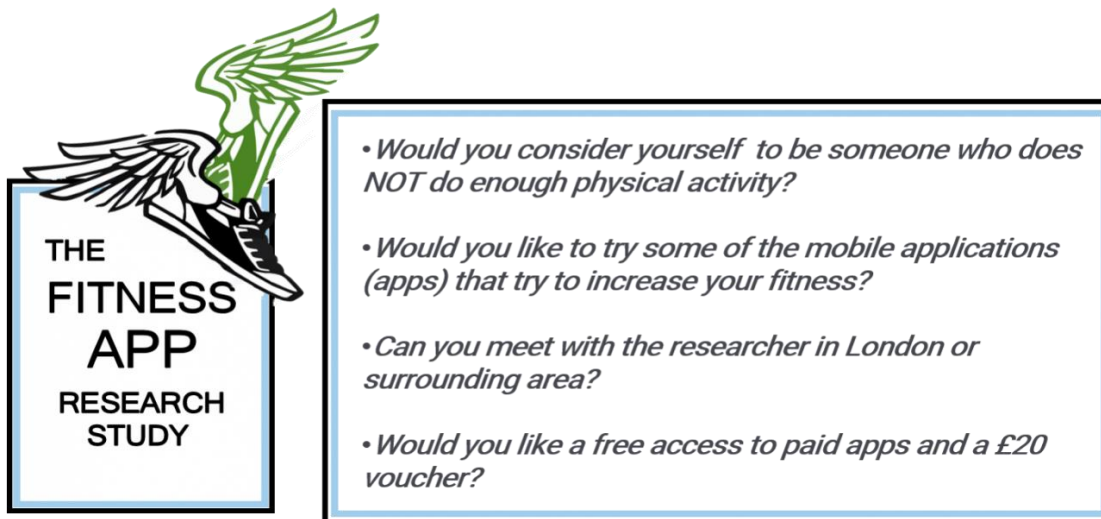
[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

[c.f.p.ac/EB6PQ3](http://c.f.p.ac/EB6PQ3)

b) screenshot of the study site on Qualtrics



**Who am I and what is the purpose of this study?**

*My name is Paulina and I am a PhD student at the University College London (UCL). I am supervised by a research team made up of doctors, researchers and lecturers at the eHealth Unit, Department of Primary Care and Population Health, UCL.*

*In this project we are assessing 2 highly popular apps that are available on the market. We would like to find out if these apps can help to increase fitness levels and what potential users (like you) think about them. We will pay for the cost of the apps.*

**What are the benefits for you?**

*You may find it interesting to use the fitness apps to see if they can help you engage in more exercise. The study will also give you free access to popular apps that you would otherwise have to pay for. Your opinions will be extremely helpful to the research and I will offer you £20 Love2shop voucher as a token of appreciation for your time and effort.*

**Who are we looking for?**

*We would like to invite adults (at least 18 years' old) who are not physically active (or do very little exercise) and who live or work in London or surrounding areas so that you can meet the researcher face-to-face at least once. You would need to own a smartphone.*



**What will you have to do?**

*First, I would like you to complete the screening questionnaire to see if you are eligible to take part. This will take around 5 minutes.*

*The study will take 5 weeks. Within this period you will be asked to try to use the apps and to make some notes on your experiences of using the apps. You will wear a small device that measures your activity and I ask you to fill in some simple survey questions. After 5 weeks there will be an interview to share with me your experiences of using the apps. In addition, I will call you/text you after 2 months to ask you if you are still using the apps. The whole process is confidential and your name will never be used in the research outputs.*

*Thank you*

*Paulina*

*If you have any questions, please contact me:*

*Email: [appapp@ucl.ac.uk](mailto:appapp@ucl.ac.uk)*

*Mobile: 07707100795*

*Data Protection Act 1998: The personal information that you give for this survey will only be used for the purpose of the survey and will not be transferred to an organisation outside of UCL.*

*By filing in the questionnaire you agree to UCL holding your personal data for just enough time to assess your eligibility. If you are not eligible, UCL will destroy any sensitive information about you.*

**Take the 5 min screening questionnaire to see if you are eligible**



Appendix M: Sensitivity analyses using vector magnitude (Study 3)

The results of the primary analysis using vector magnitude CPM displaying similar results to the ones observed using Axis 1 CPM (daily PA count).

	Period 1			p-value	Period 2			p-value
	LSM	95% CI			LSM	95% CI		
		lower	upper			lower	upper	
Daily PA count	16.58	-35.80	68.96	0.531	15.87	-42.55	74.28	0.591
MVPA	0.96	-4.43	6.36	0.723	0.91	-5.11	6.93	0.764
Light	-5.06	-11.76	1.64	0.137	-0.08	-7.55	7.40	0.983
Moderate	0.24	-4.81	5.30	0.924	1.12	-4.52	6.75	0.695
Vigorous	0.67	-0.68	2.02	0.329	-0.14	-1.65	1.37	0.855
SB	-13.81	-36.57	8.95	0.231	17.46	-7.93	42.84	0.175
Step count	73.59	-627.34	774.52	0.835	95.21	-686.55	876.97	0.809
Wear time	-17.85	-40.61	4.90	0.123	18.22	-7.16	43.61	0.157

Appendix N: Results of the sensitivity analysis of self-reported physical activity variable log transformed (Study 3)

IPAQ variable

	Period 1				Period 2			
	LSM	95% CI		p-value	LSM	95% CI		p-value
		lower	upper			lower	upper	
Total time in PA (min/week)	32.52	10.18	54.87	0.005	25.16	2.39	47.93	0.031
Moderate PA (MET-min/week)	113.68	14.97	212.38	0.024	98.49	-2.10	199.07	0.055
Vigorous PA (MET-min/week)	1.85	-201.86	205.56	0.986	37.77	-169.82	245.36	0.719
Walking (MET-min/week)	375.01	106.17	643.85	0.007	366.85	92.88	640.81	0.009
Total PA (MET-min/week)	489.46	118.18	860.75	0.010	504.22	125.86	882.59	0.010
SB (min/week)	-123.23	-196.12	-50.34	0.001	-144.77	-218.48	-71.05	0.000

Appendix O: Results of the sensitivity analysis for expected outcomes of physical activity (Study 3)

- all records (including the extreme case),
- with the outlier set to the 2nd most extreme value of -16 change in Expected Outcomes for PA

	Period 1			p-value	Period 2			p-value
	LSM	95% CI			LSM	95% CI		
		lower	upper			lower	upper	
Expected Outcomes for PA (all records)	-1.70	-3.03	-0.38	0.01	-1.81	-3.15	-0.46	0.01
Expected Outcomes for PA (outlier set to 2nd most extreme value, -16)	-1.70	-2.80	-0.61	0.00	-1.48	-2.60	-0.37	0.01

## Appendix P: Topic guide for data-prompted interviews (Study 4)

Thank you so much for your time and for coming along to talk with me today. You have agreed to take part in the study where I have asked you to use two apps that are available on the open market. This is an interview looking at your experiences of taking part in the study and using the apps. In particular, I would like to discuss how easy/burdensome were the questions I asked you to fill in, how did it feel having to wear the accelerometer device, and how did you feel about being given an app to use. I would also like to discuss using the apps, how was it using the apps from one day to another? I wanted to hear how helpful or unhelpful the apps were: their pros and cons, and how easy or difficult they were to use. Do you have any questions at all? The main thing I would like to understand is “What is it like to be a person who tried to increase their physical activity using an app?”

This interview will be audio taped so that we have an accurate record of your thoughts. Please be assured that the tapes and the transcript will be anonymised which means that you will not be identified and no one except me will have access to the transcript.

Before we start, it is important for you to know that there are no right or wrong answers. We are just exploring this topic and are really interested in hearing about your opinions. If you don't feel comfortable answering any questions, you don't have to. Do you have any questions before we start?

*[Turn on recorder]*

First, I would like to ask you some questions about your experience taking part in the study, including the measures I asked you to fill in and wearing the accelerometer device. Then, I would like to hear about your experiences of using the apps. Is this fine with you?

## **Acceptability of the interventions**

*[In this part of the interview, the representation of the participants' data will be used as prompts to guide the interview. Below I listed general areas/questions that will be discussed in this section. However, the interview will depend on the content of the data prompts]*

*[Note to self - REMEMBER: data prompts (survey responses, screenshots of the activity recoded by the apps, multimedia diaries)]*

*[First, the use of the first app used will be discussed. The questions will be repeated to discuss the 2<sup>nd</sup> app]*

Please could you open the app?

### **Intervention coherence and usability**

How much do you think you understood how the app worked?

Prompts: what worked / what did not work

How easy was the app to navigate?

How attractive was it for you to use?

### **Facilitators and barriers / likes and dislikes**

What problems have you encounter?

What did you feel about the app overall?

*Probes: Why? Can you tell me more about that?*

What did you like about it?

*Prompts: aspects/features of the apps and why? /how did it fit in your life?*

And what did you dislike?

*Prompts: aspects/features of the apps and why? / how did it fit in your life?*

What do you think are the advantages of using this app to try to increase exercise?

*Probe: Can you tell me more about that?*

### **Features of each app:**

PHE: Tailored in-app coach, timer, reminder, automatic physical activity tracker, integration with music library, mood tracking

7 MWC – videos, tailoring PA, reminders , logs, activity tracker

### **Perceived effectiveness**

Do you think it helped you increase your levels of PA?

*Probes: Why / why not?*

How useful/ not useful were this app in helping you to do more exercise?

*Probes:* Tell me more please. In what way? What particularly was useful/ not useful?

*Prompts:* what aspects/features you thought worked and why?

**What would make you use the app more? (dependent on the usage pattern)**

What in your opinion would be the best way to help people doing more physical activity (moving more?)

Do you think that apps can help to solve the problem of low physical activity (to be healthier?)

**How would you change it?**

**Adopting the apps into daily routines**

What were your experiences of day to day usage / non-usage of the apps in terms of getting to use the app in-between your usual activities?

*Probe:* You mentioned X, how difficult/easy was it for you to do this?

**Affective attitude**

How did you feel about using the app?

How did that make you feel?

*Probes:* proud/annoyed/happy/exhausted/guilty

How did you feel about using the different features of the app?

How did that make you feel?

*Probes:* proud/annoyed/happy/exhausted/guilty

**Burden**

How much effort /little effort do you think was needed from you to use the app?

**Opportunity costs**

Did you think you had to give anything up to use the app?

*Prompts:* Tell me more about that? Benefits /profits/values you had to give up?

**Wider context: physical activity – lifestyle, goals and values**

**Life priorities**

Can you tell me about your life priorities? (what do you value in life?)

How much do you want to do more physical activity?

Do benefits of doing physical activity outweigh the costs?

What problems have you encounter?

What would help you? What would motivate you? (have you had time when you stuck to something?)

Are there other things you might want to do that interfere with PA?

### **Achieving current and past physical activity goals**

What exercise goals have you set yourself in the past?

Have you achieved them?

What did you try or use to achieve your goals?

*Prompt: What worked? What didn't and why?*

What are some of the things that stood in the way of achieving your goal?

How would a perfect physical activity routine look for you?

### **Acceptability of the design, procedures and data collection methods**

#### **Acceptability of the design and procedures of the study**

What led you to take part in this study?

*Prompts: physical activity/ why did you want to increase it/overall health*

How did you feel about being given two apps to use? Would you rather have had only 1? What about having to stop using one and start using another?

*Probes: can you tell me more about this? What do you mean when you say X?*

Can you think of anything that would make more people want to take part in a similar study? Or would make it easier for those that did take part?

#### **Would you think there is a better way to try get people to try those apps?**

Would you think that there is a better way of asking you to try some apps to increase exercise?

#### **Acceptability of data collection methods**

What did you think about the questionnaires that I asked you to fill in?

*Prompts: daily questions via phone/questionnaires/questions itself*

*Prompts: easy/difficult; acceptable/annoying; timing of the reminders; frequency*

*Probes: can you tell me more about this?*

How was it for you to wear the accelerometer?

*Prompts: easy/difficult; acceptable/annoying, restricting*

Is there anything else you would like to share on the topic discussed so far?

Now, I would like to move on to discuss the experiences of using the 2 apps.

How was it to use the PACO app?



### **Closing Question**

Do you have any other comments about what we have discussed, or about the research as a whole, before we close the interview?

Is there anything you feel we haven't covered here that you feel we should give attention to?

*Prompts: personal views on increasing physical activity*

Is there anything else you would like to say or share?

*[Turn off recorder]*

Thank you for your time. Please let me know if you would like to receive a copy of the final report.

Note down the following:

- My perceptions of the person, my thoughts/emotions
- Perception of non-verbal communication
- Any notable events during the interview (my behaviour or participants' behaviour)
- Any issues, e.g., background noise

Appendix Q: Behaviour Change Wheel matrices showing the links between different components of the Wheel

**Links between COM-B and intervention functions**

	Intervention functions								
	Education	Persuasion	Incentivisation	Coercion	Training	Restriction	Environmental restructuring	Modelling	Enablement
Physical capability					█				█
Psychological capability	█				█				█
Physical opportunity					█	█	█		█
Social opportunity						█	█	█	█
Automatic motivation		█	█	█	█		█	█	█
Reflective motivation	█	█	█	█				█	

**Links between intervention functions and policy categories**

	Policy categories						
	Communication/marketing	Guidelines	Fiscal measures	Regulation	Legislation	Environmental/Social planning	Service provision
Education	█	█		█	█		█
Persuasion	█	█		█	█		█
Incentivisation	█	█	█	█	█		█
Coercion	█	█	█	█	█		█
Training		█	█	█	█		█
Restriction		█		█	█		
Environmental restructuring		█	█	█	█	█	
Modelling	█						█
Enablement		█	█	█	█	█	█

*Reference:*

- Michie, S., Atkins, L., & West, R. (2014). *The behaviour change wheel. A guide to designing interventions. 1st ed. Great Britain: Silverback Publishing, 1003-1010.*

a) Quality of Publicly Available Physical Activity Apps: Review and Content Analysis

Review

## Quality of Publicly Available Physical Activity Apps: Review and Content Analysis

Paulina Bondaronek<sup>1</sup>, BSc, MSc, MBPsS; Ghadah Alkhalidi<sup>2</sup>, MPH, PhD; April Slee<sup>3</sup>, MSc; Fiona L Hamilton<sup>1</sup>, MBBS, PhD, FFPH, MRCP; Elizabeth Murray<sup>1</sup>, PhD, FRCGP, FRCP (Edin)

<sup>1</sup>cHealth Unit, Research Department of Primary Care and Population Health, University College London, London, United Kingdom

<sup>2</sup>Community Health Sciences Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

<sup>3</sup>Research Department of Primary Care and Population Health, University College London, London, United Kingdom

**Corresponding Author:**

Paulina Bondaronek, BSc, MSc, MBPsS  
cHealth Unit  
Research Department of Primary Care and Population Health  
University College London  
Upper 3rd Floor, Royal Free Hospital  
Rowland Hill Street  
London, NW3 2PF  
United Kingdom  
Phone: 44 20 3002 878  
Email: [p.bondaronek@ucl.ac.uk](mailto:p.bondaronek@ucl.ac.uk)

### Abstract

**Background:** Within the new digital health landscape, the rise of health apps creates novel prospects for health promotion. The market is saturated with apps that aim to increase physical activity (PA). Despite the wide distribution and popularity of PA apps, there are limited data on their effectiveness, user experience, and safety of personal data.

**Objective:** The purpose of this review and content analysis was to evaluate the quality of the most popular PA apps on the market using health care quality indicators.

**Methods:** The top-ranked 400 free and paid apps from iTunes and Google Play stores were screened. Apps were included if the primary behavior targeted was PA, targeted users were adults, and the apps had stand-alone functionality. The apps were downloaded on mobile phones and assessed by 2 reviewers against the following quality assessment criteria: (1) users' data privacy and security, (2) presence of behavior change techniques (BCTs) and quality of the development and evaluation processes, and (3) user ratings and usability.

**Results:** Out of 400 apps, 156 met the inclusion criteria, of which 65 apps were randomly selected to be downloaded and assessed. Almost 30% apps (19/65) did not have privacy policy. Every app contained at least one BCT, with an average number of 7 and a maximum of 13 BCTs. All but one app had commercial affiliation, 12 consulted an expert, and none reported involving users in the app development. Only 12 of 65 apps had a peer-reviewed study connected to the app. User ratings were high, with only a quarter of the ratings falling below 4 stars. The median usability score was excellent—86.3 out of 100.

**Conclusions:** Despite the popularity of PA apps available on the commercial market, there were substantial shortcomings in the areas of data safety and likelihood of effectiveness of the apps assessed. The limited quality of the apps may represent a missed opportunity for PA promotion.

(*JMIR Mhealth Uhealth* 2018;6(3):e53) doi:[10.2196/mhealth.9069](https://doi.org/10.2196/mhealth.9069)

**KEYWORDS**

exercise; health behavior; mobile applications; health promotion; mHealth; eHealth review

## Introduction

### Background

Physical inactivity is an established independent risk factor for a range of serious health conditions including cardiovascular disease, diabetes mellitus, and cancer [1-3]. Physical activity (PA) is also associated with improved mental health [4,5]. The World Health Organization recommends 150 min of moderate or 75 min of vigorous intensity PA per week, yet 31.1% of adults globally fail to achieve this [6]. Behavior change interventions aiming to increase PA tend to have small to moderate effects, with sustainability of intervention effects not well established [7].

Within the new digital health care landscape, the rise of apps creates novel prospects for prevention opportunities and disease management [8]. Mobile health (mHealth) apps, as opposed to traditional face-to-face interventions, are more accessible [9] and provide a range of technology-enhanced features such as accelerometers, visualizations, tailored feedback, and reminders. In addition, recent data show that mobile phone access is now as high among ethnic minority groups in higher income countries as in the rest of the population [10], and the use of mobile phones is increasing steadily in older populations [11], thereby decreasing concerns about the effect of the digital divide on health inequalities. Hence, behavior change interventions delivered using mHealth apps could have the potential to reach a large proportion of the population, thus increasing the public health impact of their small effects [12].

The mHealth app industry has doubled in the last 2 years, with around 165,000 health apps available in the major app stores in 2016 [13]; many of them aiming to increase PA levels. Despite the wide distribution and popularity of health apps, many of them have been rapidly developed [14], and there is lack of evidence of their efficacy. For example, a meta-analysis published by Direito et al [15] found only 7 randomized controlled trials (RCTs) evaluating app intervention for PA and sedentary behavior. It is clearly not feasible for all PA apps to be evaluated by rigorous RCTs, and therefore, alternative methods of evaluating apps are needed. One way of assessing the likely effectiveness of apps is to assess the degree to which they use behavior change theory and adhere to PA guidelines. This research suggests that most PA apps only include a limited number of behavior change techniques (BCTs) [16-18], and they often fail to adhere to PA guidelines [19].

However, quality is about more than effectiveness, although there has been considerable debate about how exactly *app quality* should be defined, with a variety of frameworks available. Recent reviews by BinDhim et al [14] and Bardus et al [20] categorized and evaluated the methods used for quality assessment of apps. Both studies found a considerable variability in methods and measures used to review the quality of health apps. The approaches used to conceptualize and measure quality varied substantially, and the studies tended to focus on either the design quality or on the presence of evidence-based content but not both [20]. The authors called for more research to assess the quality of both design and content of health apps.

Health apps have the potential to be an important health care tool [21]; hence, health care quality indicators were considered appropriate to apply when assessing the quality of the apps. The concept of quality in health care is complex and multifaceted [22]. Maxwell [23] proposed six dimension of health care quality: accessibility (ease of access to all patient groups), relevance to the need of the community, effectiveness, equity (fairness in the distribution), acceptability, efficiency, and economy (desired health outcomes at the lowest cost). On the other hand, Donabedian [24] proposed a different categorization and argued for three crucial elements that pertain to the quality of health care: structure (facilities and health care professionals available), process (actions by which health care is provided), and outcomes (the results of the actions).

The dimensions of quality proposed by Maxwell and Donabedian were developed before the existence of mobile phones and apps and are perhaps more applicable to health care services provided at the point of need, that is, face-to-face. Potential new health care tools apps need a more concise approach, one that *High quality care for all: NHS Next Stage Review Final Report* [25] appears to provide. This report outlined the 10-year vision for the National Health Service (NHS) with strategies to improve the quality of care. In this report, high-quality health care was defined as being (1) safe, (2) effective, and (3) providing the most positive experience possible. These quality indicators are simple yet comprehensive and sufficiently flexible to apply to potential new health care tools such as PA apps.

### Objective

In this study, we focused on the most popular apps, which we defined as being in the top rankings of the two major app stores. What constitutes the algorithm that determines the app ranking is unknown. However, variables that indicate popularity such as user ratings, volume of ratings and reviews, download and install counts, usage, and uninstalls are likely to contribute to the ranking in the app stores [26]. In addition, potential users are more likely to focus on the top results and rarely examine the search results thoroughly [27]. This method of defining popularity has been used in other studies assessing apps [28-30], and it was selected to gain a representative sample of apps that are most likely to be used and to simulate the user experience of browsing the store to select a health app.

The aim of this study was to assess the quality of publicly available PA apps. Specific objectives were to assess the safety, effectiveness, and provision of the most positive experience in the most popular PA apps.

## Methods

### Study Design

This study is a review and a content analysis of the most popular, publicly available PA apps on the market. *Quality and Risk of Bias Checklist for Studies That Review Smartphone Applications* was used to ensure that methods for apps' review are adequately described [14].



**Textbox 1. Inclusion criteria.**

Apps were included if

- Their main goal was to increase physical activity
- They were targeted at healthy adults
- They had stand-alone functionality

**Textbox 2. Exclusion criteria.**

Apps were excluded if

- The app focused on multiple behaviors, as it would have been difficult to isolate the content pertaining to physical activity
- The target population was patients with a specific health condition, as these users were likely to have different needs to healthy adults
- They were sold as part of a pack ("bundle"), as it would not have been possible to assess the popularity of the individual apps in this bundle

**Sample Identification**

A sample of top-ranked 400 PA apps was obtained from the UK's versions of the iTunes and Google Play stores on October 17, 2016. As previous research indicated an association between price and inclusion of BCTs [18,31,32], both free and paid apps were included in the study. Apps' titles and descriptions from the "Health and Fitness" category in both stores (100 iTunes free + 100 iTunes paid + 100 Google Play free + 100 Google Play paid) were screened against the inclusion and exclusion criteria. (Textboxes 1 and 2)

**Sample Assessment**

From the apps identified, 65 were randomly selected for the assessment using the random number generator function in Excel (Microsoft). As the largest subset of health apps on the market (30%) [13] target PA, it was expected that a high number of apps would fulfil the inclusion criteria. We were undertaking a parallel study to assess the association between quality indicators and user rating, and the choice of n=65 was based on the power calculation for that parallel study.

The apps were downloaded onto an iPhone SE and 6 (running iPhone operating system [iOS, Apple Inc] 10.2.1 and 9.3.4 software, respectively) and Android Samsung Galaxy S6 and J5 (running 6.0.1 or 5.1.1 software, respectively) and assessed using a pro forma evaluation. Each app was left running in the background for 2 days for the assessors to explore any reminders or notifications. If two apps were identified as duplicates and there appeared to be consistency of design and content between both operating systems, the apps were assessed on an iPhone only. The sample identification and assessment was conducted independently by two reviewers (PB and GA), and any discrepancies were resolved through discussion.

**Data Extraction****Descriptive Data**

We extracted the following descriptive data from both app stores: app's name, brief description, type of PA targeted (eg, running, walking, and whole body workout), platform on which the app was available, developer's name, rank, number of ratings, cost, size, last update, and version.

**Application of Health Care Quality Indicators to Physical Activity Apps**

The methods of operationalizing the three quality indicators of safety, effectiveness, and provision of the most positive experience possible for the selected apps is described below.

**Safety of Physical Activity Apps**

For the safety indicator of health apps, privacy and security of users' data were considered. The privacy and security assessment was based on the recommendations of the Information Commissioners Office [33] and Online Trust Alliance [34]. It comprises of 8 questions evaluating the availability, accessibility of privacy policy, data gathering and sharing practices, and data security as is discussed in the privacy statement (see [Multimedia Appendix 1](#) for data privacy and security assessment).

**Likelihood of Effectiveness of Physical Activity Apps**

As research on PA app efficacy is lacking, the likelihood of effectiveness was assessed by quantifying the presence of BCTs. Furthermore, many quality assessment procedures include an evaluation of the intervention development processes [35,36]. For example, involving key stakeholders in the development process is important to produce an intervention that meets user needs and increases the likelihood of intervention implementation [37]. Hence, data on the organizational affiliation of the developer, as well as expert and user involvement in the development process was collected. In addition, any evidence of scientific evaluation was also extracted.

**Behavior Change Techniques**

The BCT taxonomy v1 [38] was used to assess the number of BCTs in each app and the frequency of each BCT in the app sample overall. The coding manual provides guidelines to investigate the presence of 93 BCTs in behavior change interventions and has been used in previous studies that aimed to characterize BCTs in health apps [16,28,39-41]. In line with the instructions, we coded each BCT as Absent, Present + (BCT present in all probability but evidence unclear), and Present ++ (BCT present beyond all reasonable doubt).

**Table 1.** The application of the health care quality indicators to physical activity apps.

Quality indicator of health care	Applying the indicator to health apps
Safety	Privacy and security of data
Effectiveness	Behavior change techniques (Michie et al [38]) Development and evaluation process: Organizational affiliation; Expert involvement; User involvement; and Evidence of scientific evaluation
Positive experience	User ratings Usability

#### Quality of Development Process and Evidence for Evaluation

The evaluation of the quality of development process was based on the information provided in the app stores, the app website (if existent), and within the app itself. The following characteristics of the app content development were extracted: organizational affiliation (university, medical, government, or other nonprofit institutions); expert involvement (eg, fitness expert, behavior change specialist, and medical professional); and evidence for user involvement in the development of an app. The evidence for app evaluation was assessed by searching the name of the app in the following scientific databases: PubMed, ACM Digital Library, IEEE Xplore, and Google Scholar.

#### Provision of the Most Positive Experience in Physical Activity Apps

The provision of the most positive experience was operationalized using (1) the user ratings in app stores and (2) through formal usability assessment conducted by the two reviewers using the System Usability Scale (SUS) [42]. The average star rating (range: 1-5 stars) was calculated by summing the number of stars and dividing them by the number of users who submitted ratings. SUS is a valid and reliable measure of overall usability (from 0-100) and consists of 10 items that are ranked on a 5-point Likert scale, from *strongly disagree* to *strongly agree*. The wording of the 8th statement was changed from *cumbersome* to *awkward* as recommended [43-45]. Second, the word *system* was replaced by *app* to make the scale applicable to the sample in this study. The interpretation of the SUS score used the thresholds proposed and validated by Bangor et al [43].

#### Summary of Application of Quality Indicators

The application of health care quality indicators to apps is summarized in Table 1.

#### Interrater Reliability

Interrater reliability for the presence or absence of the BCTs was ascertained by calculating Cohen kappa statistic [46] for each item. In addition, prevalence-adjusted bias-adjusted kappa (PABAK) [47] was assessed for the presence or absence of BCTs. The occurrence of high prevalence of negative agreement (when both rates agree that the BCT is absent) is very likely in the context of inclusion of BCTs in an app. When high prevalence of the identical response is seen, the kappa value

results in low proportion of agreement, although the observed agreement is high [48]. The a priori strategy for assessing the sample was to complete the extraction of data for 10 apps to resolve any discrepancies in understanding of the measures before extracting the rest of data. Hence, the interrater reliability was assessed on 55 apps.

#### Statistical Analysis

The number of BCTs in the apps was summarized using the mean, standard deviation, median, 25<sup>th</sup> and 75<sup>th</sup> percentiles, and the maximum and minimum. Similar statistics were used to summarize user ratings, cost, size, and SUS score. Proportions were used to summarize the variables: data privacy and security, organization affiliation, expert and user involvement, and the evidence of evaluation in peer-reviewed journals.

The summary descriptive tables were presented for each store for free and paid apps separately and in total as app stores have separate rankings based on the cost. To assess if there was a difference in store characteristics between free and paid apps, *t* tests were used to compare the average user ratings, size, and the number of BCTs; Wilcoxon test was used to compare the number of ratings; and Fisher exact was used for last update (<3 months, 3-6 months, and >6 months), organizational affiliation, expert and user involvement, and presence of any peer-reviewed studies.

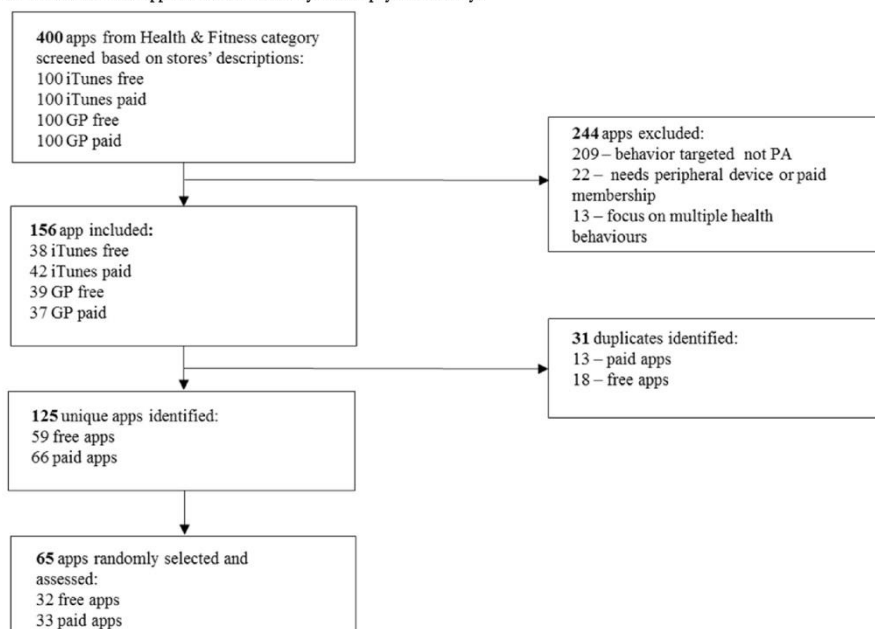
## Results

#### Sample Identification

Out of 400 apps, 244 apps were excluded (209 apps did not target PA, 22 apps needed a peripheral device or paid membership to use the app, and 13 apps focused on multiple health behaviors), and 156 met the inclusion criteria (see Figure 1). A total of 31 duplicates were found. Subsequently, a sample of 125 unique apps was identified. A total of 65 apps, 32 free and 33 paid, were assessed.

#### Sample Characteristics

Descriptive data for the app sample are presented in Tables 2 and 3, whereas the data for each app separately is presented in Multimedia Appendix 2. There were no statistically significant differences in the number of ratings, cost, size, and last update between the free and paid apps in either iTunes or Google Play store.

**Figure 1.** Flowchart of the apps included in the analysis. PA: physical activity.**Table 2.** Descriptive data for iTunes store.

Descriptive data for iTunes	Free—iTunes (N=21)	Paid—iTunes (N=24)	Total—iTunes (N=45)	P value
<b>Number of ratings</b>				
Mean (SD)	3408.4 (5848.4)	773.7 (1187.0)	2031.2 (4289.7)	.49
Median	758	127	550	
25-75 percentile	438.0-3698.0	47.0-1247.0	85.5-1719.0	
Min-max	14-24530	11-3845	11-24530	
<b>Cost—iTunes (GBP<sup>a</sup>)</b>				
Mean (SD)	N/A <sup>b</sup>	2.5 (1.5)	N/A	
Median	N/A	2.3	N/A	
25-75 percentile	N/A	1.5-3.0	N/A	
Min-max	N/A	1-8	N/A	
<b>Size of app (megabyte)</b>				
Mean (SD)	88.4 (49.8)	94.9 (75.4)	91.8 (64.1)	.74
Median	74.3	83.3	82.2	
25-75 percentile	52.0-131.0	61.7-102.0	58.1-104.0	
Min-max	11-164	9-376	9-376	
<b>Last update</b>				
<3 months, n (%)	13 (61.9)	7 (29.2)	20 (44.4)	.09
3-6 months, n (%)	3 (14.3)	7 (29.2)	10 (22.2)	

<sup>a</sup>GBP: British pound.<sup>b</sup>N/A: not applicable.<http://mhealth.jmir.org/2018/3/e53/>JMIR Mhealth Uhealth 2018 | vol. 6 | iss. 3 | e53 | p.5  
(page number not for citation purposes)



**Table 3.** Descriptive data for Google Play store.

Descriptive data for Google Play	Free—Google Play (N=21)	Paid—Google Play (N=16)	Total—Google Play (N=37)	P value
<b>Number of ratings</b>				
Mean (SD)	119000.7 (165085.0)	14457.9 (43700.8)	73793.0 (136723.2)	>.99
Median	44923	1720.5	5856	
25-75 percentile	5827.0-199596.0	384.5-6452.0	1475.0-78204.0	
Min-max	206-625077	7-177277	7-625077	
<b>Cost—Google Play (GBP<sup>a</sup>)</b>				
Mean (SD)	N/A <sup>b</sup>	3.6 (2.3)	N/A	
Median	N/A	2.7	N/A	
25-75 percentile	N/A	2.3-5.0	N/A	
Min-max	N/A	1-9	N/A	
<b>Size of app (megabyte)</b>				
Mean (SD)	28.4 (21.2)	43.4 (34.2)	34.9 (28.2)	.11
Median	26.8	31.5	29.6	
25-75 percentile	12.2-38.5	27.7-54.0	15.4-43.9	
Min-max	2-73	1-145	1-145	
<b>Last update</b>				
<3 months, n (%)	16 (76)	7 (44)	23 (62)	.12
3-6 months, n (%)	1 (5)	3 (19)	4 (11)	
>6 months, n (%)	4 (19)	6 (38)	10 (27)	

<sup>a</sup>GBP: British pound.

<sup>b</sup>N/A: not applicable.

The apps were categorized into five groups according to their primary focus. These were as follows: workout apps that demonstrate various exercises (31/65, 47%), tracking of movement apps that provide mapping of the running or walking or cycling routes (13/65, 20%), running programs that have prespecified goals reached by incremental increase in run-to-walk ratio (12/65, 18%), pedometers-based apps that count steps (6/65, 9%), and interval timers that enable the user to time their work or rest period (3/65, 4%).

### Data Privacy and Security

#### Availability and Accessibility of Privacy Policy

The privacy policy was available for 46 (70%, 46/65) apps overall. In one case, the link to the privacy policy was provided but did not work, and the app was indicated as not having a privacy policy. Of those that had privacy policy, only 4 (8%, 4/46) apps had a short form privacy and security notice that highlighted key data practices that were disclosed in detail in the full privacy policy (see Table 4). There were nine instances where the short form notice was not applicable because of the policy already being concise. Multilingual policies were rare, with only 5 apps having a policy in another language. Apps that were developed outside the United Kingdom were more likely to provide multilingual policies.

### Data Gathering and Sharing

Most of the apps (80%) reported collecting personally identifiable information. In one instance, the developer did not discuss the data gathering practices. In 34 instances (80%, 34/46), the developers stated that they share the data they gather with 3rd parties. There were two instances where the developer did not discuss data sharing practices. In many cases, the policies stated that “data shall not be shared, except for” followed by a list of exceptions that were vague and general. In these instances, the reviewers considered that the data were shared by the 3rd party.

### Data Security

Only 41% (19/46) of the apps described how the users’ data were protected. The privacy policies stated that data safety is important to their practices but did not provide information on how data security was ensured.

### The Presence of Behavior Change Techniques

There was “almost perfect” agreement between the reviewers for the coding of BCT presence or absence: PABAK=0.94, 95% CI 0.93-0.95, kappa=.78 (“substantial”), 95% CI 0.75-0.81.



**Table 4.** Data gathering, sharing and security as described in the privacy policy (within those that had the policy, N=46). Note: 29% (19/65) did not have a privacy policy available.

Data gathering, sharing, and security as described in the privacy policy	Free (N=24), n (%)	Paid (N=22), n (%)	Total (N=46), n (%)
<b>Is the privacy policy available without the need to download the app?</b>			
Yes	24 (100)	22 (100)	46 (100)
<b>Is the privacy policy available within the app?</b>			
No	13 (44)	16 (55)	29 (63)
Yes	11 (64)	6 (35)	17 (36)
<b>Is there a short form notice (in plain English) highlighting key data practices?</b>			
No	17 (70)	16 (72)	33 (71)
Yes	4 (16)	0 (0)	4 (8)
Not applicable	3 (12)	6 (27)	9 (19)
<b>Is the privacy policy available in any other languages?</b>			
No	20 (83)	21 (95)	41 (89)
Yes	4 (16)	1 (4)	5 (10)
<b>Does the app collect personally identifiable information?</b>			
No	2 (8)	6 (27)	8 (17)
Yes	21 (87)	16 (72)	37 (80)
Not specified	1 (4)	0 (0)	1 (2)
<b>Does the app share users' data with a 3rd party?</b>			
No	2 (8)	8 (36)	10 (22)
Yes	21 (87)	13 (59)	34 (74)
Not specified	1 (4)	1 (4)	2 (4)
<b>Does the app say how the users' data security is ensured? For example, encryption, authentication, and firewall</b>			
No	13 (54)	14 (63)	27 (58)
Yes	11 (45)	8 (36)	19 (41)

**Table 5.** Descriptive statistics for the inclusion of the behavior change techniques (BCTs).

Inclusion of the BCTs	Free (N=32)	Paid (N=33)	Total (N=65)	P value
<b>Total BCTs</b>				
Mean (SD)	6.6 (3.0)	7.5 (2.9)	7.0 (2.9)	.21
Median	7	8	8	
25-75 percentile	5.0-8.0	6.0-10.0	5.0-9.0	
Min-max	1-12	1-13	1-13	

The total number of BCTs for free and paid apps sample was similar (see Table 5). Every app contained at least one BCT, and the maximum number of BCTs was 12 for free and 13 for paid apps. The median number of BCTs was 7 for free and 8 for paid apps (see Multimedia Appendix 3 for the graph of the distribution of the BCTs in apps).

Figure 2 shows the frequency of the common BCT groups. The "Feedback and monitoring" group was the most common, with 92.3% of apps containing at least one BCT of this group, most commonly "Feedback on behavior" and "Feedback on outcome(s) of behavior" BCTs. "Goals and planning" ("Goal

setting" and "Action planning" BCTs) were also well represented at 84.6%. More than half of the apps included BCTs from the "Comparison of behavior" group (66.2%), which most likely was "Demonstration on the behavior" (see Figure 3 for the examples of the app features that included BCTs from the most common BCT groups). "Social support" (64.6%), "Shaping knowledge" groups (60%), and "Associations" (46.2%) were common, but only one BCT from each of these groups were present. "Reward and threat" group (53.8%) was common with two BCTs only ("Social reward" and "Nonspecific incentive"). Other BCT groups were rare: less than 15% of apps contained

BCTs from the “Comparison of outcomes” group; “Natural consequences” and “Antecedents” represented 10.8% and 6.2% of the total BCTs, respectively. The remaining BCT groups were nonexistent in the PA apps. **Multimedia Appendix 4** presents the frequency of individual BCTs within the groups’ BCTs (BCTs that occurred in at least five apps are shown).

**Quality of App Development and Evaluation Process**

Only 1 app had a noncommercial affiliation, *One You Couch to 5K*, which was developed by Public Health England (see **Table 6**). None of the apps reported user involvement during development. Twelve out of 65 apps (4 free and 8 paid) consulted with experts to design the content of the app. Nine out of 23 free apps (28.1%) had a study associated with the apps published in a peer-reviewed journal. In comparison, for only 3 paid apps (9.1%), there was a peer-reviewed study found.

**Positive Experience**

**User Ratings**

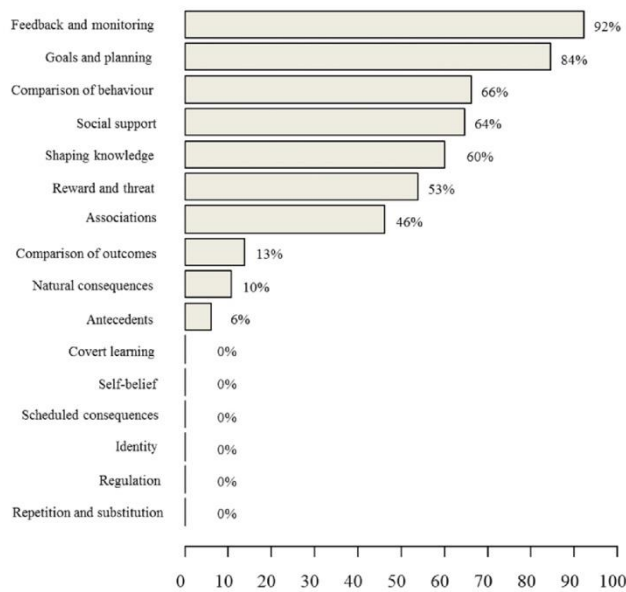
The median user rating in iTunes was 4.4 and 4.5 in Google Play and did not differ between free and paid apps in either stores (see **Table 7**).

In both stores, the 25th percentile was around 4 stars (4.0 in iTunes and 4.4 in Google Play), suggesting that the user ratings tended to be high, and only 25% of ratings were below 4 stars. The histograms of star ratings in both stores (**Figure 4**) showed the skewness of the star average distribution.

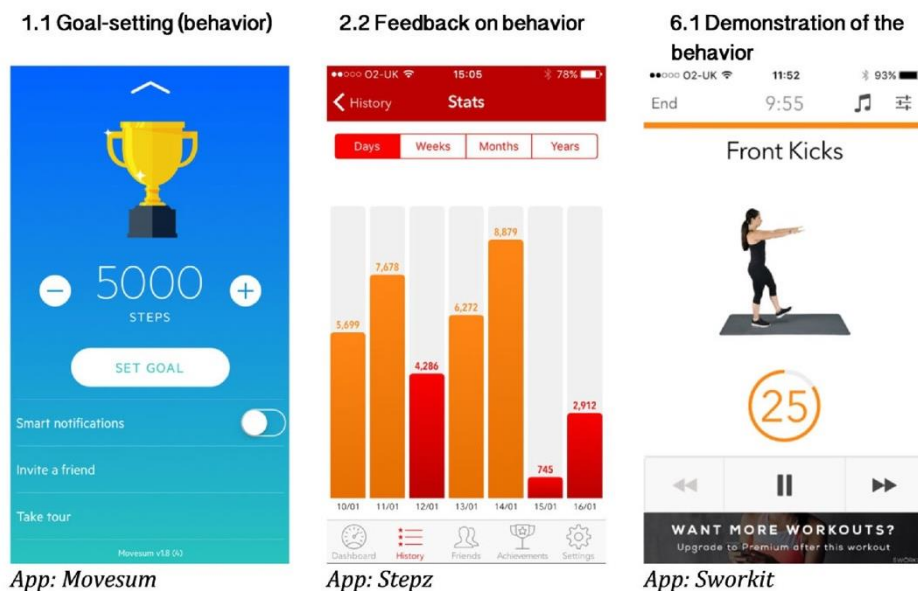
**Usability**

The average SUS score for the apps was similar for both free and paid apps, with median of 86.3 (see **Table 8**). Using the descriptors suggested by Bangor et al [43], the score can be described as “excellent.” Fifty percent of the total average SUS score fell between 75.0 and 92.5, and 25% had a score higher than 92.5, suggesting that more than 75% of the app sample assessed could be described as having “good” to “excellent” usability. See **Multimedia Appendix 5** for the graph of the distribution of the SUS score averaged between the two reviewers.

**Figure 2.** Frequency of behavior change techniques (BCTs) incorporated by physical activity (PA) apps, presented by BCT groups.



**Figure 3.** Examples of the most common behavior change techniques (BCTs) from the most frequent BCT groups: (1) goals and planning: 1.1 Goal setting (behavior), (2) feedback and monitoring: 2.2 Feedback on behavior, and (3) comparison of behavior: 6.1 Demonstration of the behavior.



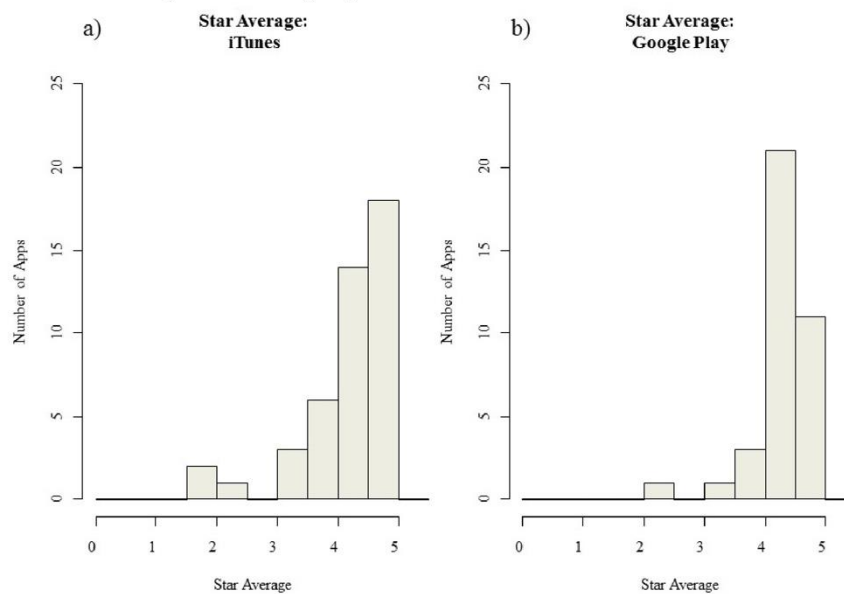
**Table 6.** Descriptive data for the quality of app development and evaluation process: organizational affiliation, expert and user involvement, and evidence of evaluation in peer-reviewed journals.

The quality of app development and evaluation process	Free (N=32), n (%)	Paid (N=33), n (%)	Total (N=65), n (%)	P value
<b>Any affiliation</b>				
Commercial	31 (96)	33 (100)	64 (98)	.49
Government institution	1 (3)	0 (0)	1 (1)	
<b>Any expert</b>				
No	28 (87)	25 (75)	53 (81)	.34
Yes	4 (12)	8 (24)	12 (18)	
<b>Any user involvement</b>				
No	32 (100)	33 (100)	65 (100)	
<b>Any peer journal</b>				
No	23 (71)	30 (90)	53 (81)	.06
Yes	9 (28)	3 (9)	12 (18)	

**Table 7.** Descriptive statistics for user ratings (1-5 stars) in iTunes and Google Play.

User ratings	Free	Paid	Total	P value
<b>iTunes</b>	(N=21)	(N=24)	(N=45)	
Mean (SD)	4.1 (0.8)	4.3 (0.6)	4.2 (0.7)	.22
Median	4.4	4.6	4.4	
25-75 percentile	4.0-4.6	4.0-4.8	4.0-4.6	
Min-max	2-5	3-5	2-5	
<b>Google Play</b>	(N=21)	(N=16)	(N=37)	
Mean (SD)	4.4 (0.5)	4.4 (0.3)	4.4 (0.4)	.90
Median	4.5	4.5	4.5	
25-75 percentile	4.4-4.6	4.4-4.6	4.4-4.6	
Min-max	2-5	4-5	2-5	

**Figure 4.** Distribution of user ratings in iTunes and Google Play.



**Table 8.** Descriptive data for the System Usability Scale (SUS) assessment.

Usability assessment	Free (N=32)	Paid (N=33)	Total (N=65)	P value
<b>SUS score</b>				.17
Mean (SD)	81.3 (12.6)	85.5 (11.9)	83.4 (12.4)	
Median	85	87.5	86.3	
25-75 percentile	71.9-91.3	80.0-93.8	75.0-92.5	
Min-max	53-100	58-100	53-100	



## Discussion

### Principal Findings

This study described the most popular PA apps on the market, focusing on the quality determinants of safety (data privacy and security), effectiveness (BCTs and development and evaluation quality), and provision of the most positive experience possible (user ratings and usability). Overall, our findings suggest that most of the apps in this sample were of reasonable quality in terms of the user experience, but there were substantial shortcomings in the areas of safety and effectiveness. The assessment of data privacy and security showed that the privacy policy was not available for 29.2% of the apps. Most apps collected personally identifiable information, shared users' data with a third party, and more than half of the apps did not specify how they ensure data security. Every app contained at least one BCT, with an average of 7. The maximum number of BCTs was 13, and the most common BCTs related to provision of feedback on behavior. All but one app had commercial affiliation, 12 consulted an expert, and none reported involving users in the app development. Only 12 of 65 apps had a peer-reviewed study connected to the app but only one app was assessed for efficacy in a trial [49]. User ratings were high, with only a quarter of the ratings falling below 4 stars. Similarly, the usability scores were "good" to "excellent." There was no statistically significant difference between free and paid apps on the characteristics or quality indicators.

### Safety of Apps

The assessment of privacy policy showed that privacy and security of users' data could be substantially improved. Our results are consistent with previous studies assessing data safety. Huckvale et al [8], who assessed the apps from the NHS Apps Library, found that 20% of apps did not have privacy policy, and most of the apps breached users' data privacy and security. Collecting and analyzing consumer data by app developers can have advantages for the users, such as personalization and improvement of the products [35]. However, the information about these practices ought to be transparent and understandable [36] to enable the potential user to make an informed decision to download the app. Regulatory oversight concerning data protection is challenging because of the large scale of the app market. In consequence, ensuring the privacy and security of data is left in the hands of app developers [50].

### Likelihood of Effectiveness

The apps in the review contained, on average, 7 BCTs. The results of this study are similar to those found in previous reviews of PA apps: Middelweerd et al [17] found that, on average, 5 BCTs were used in each app; Conroy et al [16] reported between 1 and 13 BCTs with a mean of 4.2; and a study using the same BCT taxonomy as the one in this study found, on average, 6.6 BCTs [18].

The most common BCTs were feedback and monitoring, goal setting, and action planning. These self-regulation strategies have been shown to be effective in increasing PA behavior [51,52]. However, the BCTs from 9 out of 16 BCT groups were

rare or nonexistent in the apps assessed, and the BCTs that were present constituted 14% of the current BCT taxonomy.

The effect of the number of BCTs on efficacy of the interventions remains inconclusive. Although there is some evidence that higher number of BCTs produces larger effect sizes in Web-based interventions [53], others show no effect [51]. The evidence of what BCTs are most likely to increase the likelihood of behavior change is unknown. It is possible that certain BCTs are more efficacious when present together producing a synergistic effect [54]. The use of variety of BCTs groups, as well as the techniques within the BCT group, would theoretically increase effectiveness by addressing various barriers to PA. For example, within the "Goals and planning" BCT group, only 3 out of 9 BCTs were utilized. Implementing features that utilize other BCTs that enable goal setting and planning (eg, problem-solving technique, asking the user to commit to their goal, and providing an opportunity for the user to review their goal) might increase the likelihood of effectiveness of the app.

The use of evidence and theoretical frameworks is vital in developing behavior change interventions [55]. The COM-B (capability, opportunity, motivation, and *behavior*) model of behavior change [56] enables developers to systematically identify the barriers and facilitators of the behavior targeted and to select intervention components that will address these barriers to increase the likelihood of behavior change.

The results suggest that the quality of the app development and evaluation process could be improved. We did not find any evidence of user involvement, and most apps were commercially developed with the rare involvement of experts. Similar results were found in previous reviews [28,57], and there is evidence to suggest that expert involvement predict the number of app download [58]. Indeed, the user-centered design framework stresses the importance of understanding the contextual experiences of potential users, as well as inclusion of multidisciplinary skills and perspectives when developing products and services. Our results also support previous research showing the lack of evidence for scientific evaluation of the apps on the market [59,60]. We found only 12 studies in peer-reviewed journals that were associated with the apps. However, only one app was used in a pragmatic RCT [49], and the study was not conducted by the app developer.

### Positive Experience

The usability of the apps reviewed was high. Likewise, user ratings of the PA apps were high, with only a quarter of the ratings receiving less than 4 stars. Similarly, Mendiola et al [61] found that usability was related to user ratings in a general sample of health apps. The competition for customer in the app stores is high, with 90% of apps in the app stores not attracting enough attention to feature in the ranking of the app stores and consequently not visible for the user, called "App Zombies" [62]. High-quality graphic design, visual appeal, and ease of use are more likely to attract potential customers to download and engage with the app. However, it is unknown whether these variables relate to effectiveness of the apps. There is evidence to suggest that Web-based interventions with higher usability tend to be more effective [54]. However, continued engagement

with an app may suggest engagement with the intervention or unhealthy dependence [63].

### Strengths

The strengths of this study include a systematic approach to sample identification and assessment. First, the sample of apps was identified by screening 400 apps in two major app distribution platforms, including both free and paid apps. Second, the sample was identified and assessed by 2 independent reviewers. Third, the assessment tools covered various aspects of quality, both inclusion of theory as well as user experience using subjective (user ratings) and objective (usability) measures.

### Limitations

First, it is unknown what variables are included in the ranking algorithm of the top apps from which the sample was selected. It is likely that usage data and user ratings comprise the ranking [26], but other unknown variables may also be included. Second, the possibility that user ratings were influenced by fake reviews cannot be excluded. [64,65]. However, there is a reliance on genuine users of the app to mark it down if the app does not live to their expectations, and this review included popular apps with high number of ratings (2.8 million). Third, data privacy and security assessment was limited to the analysis of the policy. There is evidence of inconsistency between the policy statement and the actual practices of app developers [8]. Fourth, the quality of app development process was based on the information provided in the app stores, the app website, and within the app itself; hence, it is possible that some data were missed if they were not available on the Web. Finally, the evidence for app evaluation was assessed by searching the name of the app in the popular scientific databases. If the name of the app was absent in the title or abstract, then the relevant paper would not have been found.

### Implications

More studies are needed to assess what predicts higher user rating. It is unknown what features or characteristics of apps

users like and perceive to be effective in increasing their PA. It is possible that there is a discrepancy between what is liked and what is more likely to be effective. Second, research is needed to understand the use of PA apps to design effective digital tools. There is little knowledge concerning how users adopt these apps into their routines and what are the facilitators and barriers to increasing PA using apps. Third, the optimal number of BCTs in PA app remains unknown. It is likely that different BCTs may be more suitable for different modes of delivery (face-to-face, Web-based, and app). For example, social support might produce better results when delivered face-to-face rather than via an app. Alternatively, automatic monitoring and feedback on PA in apps can facilitate self-regulation and may be considered as a more efficient method than self-monitoring using diaries.

Although popularity of the apps is high, health care professionals and potential users need to be aware of the limitation in the safety of personal data, as well as the limitation in the quality of the apps to change behavior. Currently, it is not possible to recommend apps that are most effective, but attempts to create a database of high-quality apps are in progress. For example, the National Information Board is developing an app accreditation model that consists of a 4-stage assessment framework that aims to establish a database of high-quality health apps [66].

### Conclusions

This study examined the quality of the most popular PA apps currently available on the market. Although usability and user ratings of app were high, there was a concerning lack of safety controls for users' personal data for the majority of the apps, the apps included limited number of BCTs that mostly related to feedback on behavior, and the quality of the content and development processes were suboptimal. The technological development and the potential for profit far outpaced the research on the ability of these apps to support PA behavior change. With 165,000 apps on the market, this represents a loss of opportunity for health promotion on a large scale.

---

### Acknowledgments

The authors would like to thank Lou Atkins, Senior Teaching Fellow, Department of Clinical, Educational and Health Psychology, University College London, London, UK, who was consulted on the BCT coding. PK is a PhD student at University College London, funded by the Medical Research Council.

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Data privacy and security assessment based in the content of privacy policy.

[PDF File (Adobe PDF File), 99KB - [mhealth\\_v6i3e53\\_app1.pdf](#) ]

---

### Multimedia Appendix 2

Individual-level data for the sample of apps assessed.

[PDF File (Adobe PDF File), 649KB - [mhealth\\_v6i3e53\\_app2.pdf](#) ]

---

<http://mhealth.jmir.org/2018/3/e53/>

JMIR Mhealth Uhealth 2018 | vol. 6 | iss. 3 | e53 | p.12  
(page number not for citation purposes)



**Multimedia Appendix 3**

Graph of the distribution of the BCTs in PA apps.

[JPG File, 52KB - [mhealth\\_v6i3e53\\_app3.jpg](#)]

**Multimedia Appendix 4**

Frequency of individual BCTs within the groups BCTs (BCTs that occurred in at least five apps are shown).

[JPG File, 217KB - [mhealth\\_v6i3e53\\_app4.jpg](#)]

**Multimedia Appendix 5**

Graph of the distribution of the SUS score averaged between the two reviewers.

[PDF File (Adobe PDF File), 101KB - [mhealth\\_v6i3e53\\_app5.pdf](#)]

**References**

1. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports* 2015 Dec;25(Suppl 3):1-72. [doi: [10.1111/sms.12581](#)] [Medline: [26606383](#)]
2. Knols R, Aaronson NK, Uebelhart D, Fransen J, Aufdemkampe G. Physical exercise in cancer patients during and after medical treatment: a systematic review of randomized and controlled clinical trials. *J Clin Oncol* 2005 Jun 01;23(16):3830-3842. [doi: [10.1200/JCO.2005.02.148](#)] [Medline: [15923576](#)]
3. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. *Compr Physiol* 2012 Apr;2(2):1143-1211 [FREE Full text] [doi: [10.1002/cphy.c110025](#)] [Medline: [23798298](#)]
4. Chen JJ, Lee Y. Physical activity for health: evidence, theory, and practice. *J Prev Med Public Health* 2013 Jan;46(Suppl 1):S1-S2 [FREE Full text] [doi: [10.3961/jpmph.2013.46.S.S1](#)] [Medline: [23413397](#)]
5. Zschucke E, Gaudlitz K, Ströhle A. Exercise and physical activity in mental disorders: clinical and experimental evidence. *J Prev Med Public Health* 2013 Jan;46(Suppl 1):S12-S21 [FREE Full text] [doi: [10.3961/jpmph.2013.46.S.S12](#)] [Medline: [23412549](#)]
6. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet* 2012 Jul 21;380(9838):247-257. [doi: [10.1016/S0140-6736\(12\)60646-1](#)] [Medline: [22818937](#)]
7. Dzawaltowski DA, Estabrooks PA, Glasgow RE. The future of physical activity behavior change research: what is needed to improve translation of research into health promotion practice? *Exerc Sport Sci Rev* 2004 Apr;32(2):57-63. [Medline: [15064649](#)]
8. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Med* 2015;13:214 [FREE Full text] [doi: [10.1186/s12916-015-0444-y](#)] [Medline: [26404673](#)]
9. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. *Clin Psychol Psychother* 2014;21(5):427-436 [FREE Full text] [doi: [10.1002/cpp.1855](#)] [Medline: [23918764](#)]
10. Pew Internet. African Americans and technology use URL: <http://www.pewinternet.org/files/2014/01/African-Americans-and-Technology-Use.pdf> [accessed 2016-05-05] [WebCite Cache ID 6tk0yhui5]
11. Office for National Statistics. 2017. Internet access—households and individuals URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017> [accessed 2017-10-01] [WebCite Cache ID 6ualcz5xB]
12. Glasgow RE, Bull SS, Piette JD, Steiner JF. Interactive behavior change technology: a partial solution to the competing demands of primary care. *Am J Prev Med* 2004 Aug;27(2 Suppl):80-87. [doi: [10.1016/j.amepre.2004.04.026](#)] [Medline: [15275676](#)]
13. Research2Guidance. 2014. mHealth app developer economics 2014 URL: <http://research2guidance.com/product/mhealth-app-developer-economics-2014/> [accessed 2017-09-25] [WebCite Cache ID 6tk2FAJel]
14. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. *Telemed J E Health* 2015 Feb;21(2):97-104. [doi: [10.1089/tmj.2014.0088](#)] [Medline: [25469795](#)]
15. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. *Ann Behav Med* 2017 Apr;51(2):226-239. [doi: [10.1007/s12160-016-9846-0](#)] [Medline: [27757789](#)]
16. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. *Am J Prev Med* 2014 Jun;46(6):649-652. [doi: [10.1016/j.amepre.2014.01.010](#)] [Medline: [24842742](#)]

17. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. *Int J Behav Nutr Phys Act* 2014;11:97 [FREE Full text] [doi: [10.1186/s12966-014-0097-9](https://doi.org/10.1186/s12966-014-0097-9)] [Medline: [25059981](#)]
18. Yang C, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. *Am J Prev Med* 2015 Apr;48(4):452-455. [doi: [10.1016/j.amepre.2014.10.010](https://doi.org/10.1016/j.amepre.2014.10.010)] [Medline: [25576494](#)]
19. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. *JMIR Mhealth Uhealth* 2015;3(2):e43 [FREE Full text] [doi: [10.2196/mhealth.4003](https://doi.org/10.2196/mhealth.4003)] [Medline: [25998158](#)]
20. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and web 2.0 technologies for weight management: a systematic scoping review. *J Med Internet Res* 2015;17(11):e259 [FREE Full text] [doi: [10.2196/jmir.5129](https://doi.org/10.2196/jmir.5129)] [Medline: [26573984](#)]
21. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G, et al. Mobile phones as medical devices in mental disorder treatment: an overview. *Pers Ubiquitous Comput* 2014 Sep 21;19(2):335-353. [doi: [10.1007/s00779-014-0829-5](https://doi.org/10.1007/s00779-014-0829-5)]
22. Campbell SM, Roland MO, Buetow SA. Defining quality of care. *Soc Sci Med* 2000 Dec;51(11):1611-1625. [Medline: [11072882](#)]
23. Maxwell RJ. Quality assessment in health. *Br Med J (Clin Res Ed)* 1984 May 12;288(6428):1470-1472 [FREE Full text] [Medline: [6426606](#)]
24. Donabedian A. The quality of care. How can it be assessed? *J Am Med Assoc* 1988;260(12):1743-1748. [Medline: [3045356](#)]
25. Department of Health. 2008. High quality care for all: NHS next stage review final report URL: <https://www.gov.uk/government/publications/high-quality-care-for-all-nhs-next-stage-review-final-report> [accessed 2017-10-01] [WebCite Cache ID [6tk2SKca8](#)]
26. Waltz A. Moz. 2015 May 27. Deconstructing the app store rankings formula with a little mad science URL: <https://moz.com/blog/app-store-rankings-formula-deconstructed-in-5-mad-science-experiments> [accessed 2018-02-15] [WebCite Cache ID [6xFD9TS31](#)]
27. Infront. Value of organic first-page results 2015 URL: <https://www.infront.com/blogs/the-infront-blog/2015/6/17/value-of-first-page-google-results> [accessed 2017-09-25] [WebCite Cache ID [6tk3VmF4i](#)]
28. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. *JMIR Mhealth Uhealth* 2015;3(4):e104 [FREE Full text] [doi: [10.2196/mhealth.4334](https://doi.org/10.2196/mhealth.4334)] [Medline: [26678569](#)]
29. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. *Am J Prev Med* 2013 Nov;45(5):583-589. [doi: [10.1016/j.amepre.2013.07.005](https://doi.org/10.1016/j.amepre.2013.07.005)] [Medline: [24139771](#)]
30. Breton ER, Fuemmeler BF, Abrams LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? *Transl Behav Med* 2011 Dec;1(4):523-529 [FREE Full text] [doi: [10.1007/s13142-011-0076-5](https://doi.org/10.1007/s13142-011-0076-5)] [Medline: [24073074](#)]
31. Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: a content analysis of exercise apps for presence of behavior change theory. *Health Educ Behav* 2013 Apr;40(2):133-139. [doi: [10.1177/1090198112452126](https://doi.org/10.1177/1090198112452126)] [Medline: [22991048](#)]
32. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. *J Med Internet Res* 2012;14(3):e72 [FREE Full text] [doi: [10.2196/jmir.1977](https://doi.org/10.2196/jmir.1977)] [Medline: [22584372](#)]
33. Information Commissioner's Office. 2013. Privacy in mobile apps URL: <https://ico.org.uk/media/for-organisations/documents/1596/privacy-in-mobile-apps-dp-guidance.pdf> [accessed 2017-09-25] [WebCite Cache ID [6tk3cjdFA](#)]
34. Online Trust Alliance. 2014. Mobile App Privacy & Security URL: <https://otalliance.org/best-practices/mobile-app-privacy-security> [accessed 2018-02-14] [WebCite Cache ID [6xEYd7pet](#)]
35. National Institute for HealthCare Excellence. NICE. 2014 Jan 02. Behavior change: individual approaches URL: <https://www.nice.org.uk/guidance/ph49/chapter/What-is-this-guidance-about> [accessed 2017-10-01] [WebCite Cache ID [6tk3dtx4L](#)]
36. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. *Issue Brief (Commonw Fund)* 2016 Feb;5:1-11. [Medline: [26934758](#)]
37. 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;337:a1655 [FREE Full text] [Medline: [18824488](#)]
38. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013 Aug;46(1):81-95. [doi: [10.1007/s12160-013-9486-6](https://doi.org/10.1007/s12160-013-9486-6)] [Medline: [23512568](#)]
39. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. *J Med Internet Res* 2015;17(5):e118 [FREE Full text] [doi: [10.2196/jmir.4060](https://doi.org/10.2196/jmir.4060)] [Medline: [25977135](#)]
40. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *Int J Behav Nutr Phys Act* 2016;13(1):35 [FREE Full text] [doi: [10.1186/s12966-016-0359-9](https://doi.org/10.1186/s12966-016-0359-9)] [Medline: [26964880](#)]



41. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? *BMC Public Health* 2014;14:646 [FREE Full text] [doi: [10.1186/1471-2458-14-646](https://doi.org/10.1186/1471-2458-14-646)] [Medline: [24965805](https://pubmed.ncbi.nlm.nih.gov/24965805/)]
42. Brooke J. Hell.meiert. Reading, UK: Digital Equipment Co Ltd; 1986. System usability scale (SUS): a quick-and-dirty method of system evaluation user information URL: <https://hell.meiert.org/core/pdf/sus.pdf> [accessed 2018-02-15] [WebCite Cache ID 6xFl7FGOo]
43. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum Comput Interact* 2008 Jul 30;24(6):574-594. [doi: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776)]
44. Finstad K. The system usability scale and non-native english speakers. *J Usability Stud* 2006;1(4):185-188.
45. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int J Hum Comput Interact* 2011 Jun;14(3-4):463-488. [doi: [10.1080/10447318.2002.9669130](https://doi.org/10.1080/10447318.2002.9669130)]
46. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica* 2012;22(3):276-282. [doi: [10.11613/BM.2012.031](https://doi.org/10.11613/BM.2012.031)]
47. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993 May;46(5):423-429. [doi: [10.1016/0895-4356\(93\)90018-V](https://doi.org/10.1016/0895-4356(93)90018-V)]
48. Cunningham M. SAS. 2009. More than just the kappa coefficient: a program to fully characterize inter-rater reliability between two raters URL: <http://support.sas.com/resources/papers/proceedings09/242-2009.pdf> [accessed 2018-02-15] [WebCite Cache ID 6xFljlmYO]
49. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for improving FITness and increasing physical activity among young People: the AIMFIT pragmatic randomized controlled trial. *J Med Internet Res* 2015;17(8):e210 [FREE Full text] [doi: [10.2196/jmir.4568](https://doi.org/10.2196/jmir.4568)] [Medline: [26316499](https://pubmed.ncbi.nlm.nih.gov/26316499/)]
50. Wicks P, Chiauzzi E. 'Trust but verify'--five approaches to ensure safe medical apps. *BMC Med* 2015;13:205 [FREE Full text] [doi: [10.1186/s12916-015-0451-z](https://doi.org/10.1186/s12916-015-0451-z)] [Medline: [26404791](https://pubmed.ncbi.nlm.nih.gov/26404791/)]
51. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychol* 2009 Nov;28(6):690-701. [doi: [10.1037/a0016136](https://doi.org/10.1037/a0016136)] [Medline: [19916637](https://pubmed.ncbi.nlm.nih.gov/19916637/)]
52. Pearson ES. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. *Patient Educ Couns* 2012 Apr;87(1):32-42. [doi: [10.1016/j.pec.2011.07.018](https://doi.org/10.1016/j.pec.2011.07.018)] [Medline: [21852063](https://pubmed.ncbi.nlm.nih.gov/21852063/)]
53. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *J Med Internet Res* 2010;12(1):e4 [FREE Full text] [doi: [10.2196/jmir.1376](https://doi.org/10.2196/jmir.1376)] [Medline: [20164043](https://pubmed.ncbi.nlm.nih.gov/20164043/)]
54. van Genugten L, Dusseldorp E, Webb TL, van Empelen EP. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. *J Med Internet Res* 2016;18(6):e155 [FREE Full text] [doi: [10.2196/jmir.4218](https://doi.org/10.2196/jmir.4218)] [Medline: [27268104](https://pubmed.ncbi.nlm.nih.gov/27268104/)]
55. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Annu Rev Public Health* 2010 Mar;31:399-418. [doi: [10.1146/annurev.publhealth.012809.103604](https://doi.org/10.1146/annurev.publhealth.012809.103604)] [Medline: [20070207](https://pubmed.ncbi.nlm.nih.gov/20070207/)]
56. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011;6:42 [FREE Full text] [doi: [10.1186/1748-5908-6-42](https://doi.org/10.1186/1748-5908-6-42)] [Medline: [21513547](https://pubmed.ncbi.nlm.nih.gov/21513547/)]
57. Subhi Y, Bube SH, Rolskov BS, Skou TAS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. *JMIR Mhealth Uhealth* 2015 Jul 27;3(3):e79 [FREE Full text] [doi: [10.2196/mhealth.4169](https://doi.org/10.2196/mhealth.4169)] [Medline: [26215371](https://pubmed.ncbi.nlm.nih.gov/26215371/)]
58. Pereira-Azcvedo N, Osório L, Cavadas V, Fraga A, Carrasquinho E, Cardoso de Oliveira E, et al. Expert involvement predicts mHealth app downloads: multivariate regression analysis of urology apps. *JMIR Mhealth Uhealth* 2016;4(3):e86 [FREE Full text] [doi: [10.2196/mhealth.5738](https://doi.org/10.2196/mhealth.5738)] [Medline: [27421338](https://pubmed.ncbi.nlm.nih.gov/27421338/)]
59. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. *J Med Internet Res* 2013;15(6):e120 [FREE Full text] [doi: [10.2196/jmir.2600](https://doi.org/10.2196/jmir.2600)] [Medline: [23770578](https://pubmed.ncbi.nlm.nih.gov/23770578/)]
60. Modave F, Bian J, Leavitt T, Bromwell J, Harris IC, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine guidelines: a review of current mobile apps. *JMIR Mhealth Uhealth* 2015;3(3):e77 [FREE Full text] [doi: [10.2196/mhealth.4669](https://doi.org/10.2196/mhealth.4669)] [Medline: [26209109](https://pubmed.ncbi.nlm.nih.gov/26209109/)]
61. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. *JMIR Mhealth Uhealth* 2015 May 13;3(2):e40 [FREE Full text] [doi: [10.2196/mhealth.4283](https://doi.org/10.2196/mhealth.4283)] [Medline: [25972309](https://pubmed.ncbi.nlm.nih.gov/25972309/)]
62. Adjust. 2016. The Zombie Uprising: a look at the undead App Store in 2016 URL: <https://www.adjust.com/resources/insights/app-zombies-2016/> [accessed 2018-02-14] [WebCite Cache ID 6xEWc7zgL]
63. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
64. Girardello A, Michahelles F. AppAware: which mobile applications are hot? 2010 Sep Presented at: Proceedings of the 12th International Conference on Human Computer Interaction With Mobile Devices Services; 2010; Lisbon, Portugal.

65. Kuehnhausen M, Frost VS. Trusting smartphone apps? To install or not to install, that is the question. 2013 Feb Presented at: 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA); 2013; San Diego, CA, USA. [doi: [10.1109/CogSIMA.2013.6523820](https://doi.org/10.1109/CogSIMA.2013.6523820)]
66. Gov.uk. Personalised health Care 2020: work stream 1.2 Roadmap. Enable me to make the right health and care choices. Providing citizens with access to an assessed set of NHS and social care 'apps' URL: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/433170/NIB\\_WS\\_1\\_2.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/433170/NIB_WS_1_2.pdf) [accessed 2017-09-25] [WebCite Cache ID 6tk3fDkZP]

#### Abbreviations

**BCT:** behavior change technique  
**mHealth:** mobile health  
**NHS:** National Health Service  
**PA:** physical activity  
**PABAK:** prevalence-adjusted bias-adjusted kappa  
**RCT:** randomized controlled trial  
**SUS:** System Usability Scale

*Edited by G Eysenbach; submitted 05.10.17; peer-reviewed by J Lumsden, H Atherton; comments to author 25.10.17; revised version received 23.11.17; accepted 23.11.17; published 21.03.18*

*Please cite as:*

Bondaronck P, Alkhalidi G, Slec A, Hamilton FL, Murray E  
Quality of Publicly Available Physical Activity Apps: Review and Content Analysis  
JMIR Mhealth Uhealth 2018;6(3):e53  
URL: <http://mhealth.jmir.org/2018/3/e53/>  
doi: [10.2196/mhealth.9069](https://doi.org/10.2196/mhealth.9069)  
PMID: [29563080](https://pubmed.ncbi.nlm.nih.gov/29563080/)

©Paulina Bondaronck, Ghadah Alkhalidi, April Slec, Fiona L Hamilton, Elizabeth Murray. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 21.03.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.



## b) Relationship between popularity and the likely efficacy: an observational study based on a random selection on top-ranked physical activity apps

Open access

Original research

# BMJ Open Relationship between popularity and the likely efficacy: an observational study based on a random selection on top-ranked physical activity apps

Paulina Bondaronek <sup>1</sup>, April Slee,<sup>2</sup> Fiona L Hamilton,<sup>1</sup> Elizabeth Murray <sup>1</sup>

**To cite:** Bondaronek P, Slee A, Hamilton FL, *et al.* Relationship between popularity and the likely efficacy: an observational study based on a random selection on top-ranked physical activity apps. *BMJ Open* 2019;**9**:e027536. doi:10.1136/bmjopen-2018-027536

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-027536>).

Received 31 October 2018  
Revised 24 September 2019  
Accepted 26 September 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY. Published by BMJ.

<sup>1</sup>eHealth Unit, Research Department of Primary Care and Population Health, University College London, London, UK  
<sup>2</sup>Research Department of Primary Care and Population Health, University College London, London, UK

**Correspondence to**  
Dr Paulina Bondaronek;  
[p.bondaronek@ucl.ac.uk](mailto:p.bondaronek@ucl.ac.uk)

### ABSTRACT

**Objectives** To explore the relationship between popularity of mobile application (apps) for physical activity (PA) and their likely efficacy. The primary objective was to assess the association between app popularity (indicated by user ratings) and likely efficacy (indicated by the number of Behaviour Change Techniques (BCT) present). The secondary objective was to assess the relationship between user ratings and those BCTs that have been shown to be effective in increasing PA.

**Design** Observational study.

**Methods** 400 top-ranked free and paid apps from iTunes and Google Play stores were screened, and were included if the primary behaviour targeted was PA and they had stand-alone functionality. The outcome variable of user rating was dichotomised into high (4, 5 stars) or low (1, 2, 3 stars) rating.

**Setting** iTunes and Google Play app stores.

**Participants** No individual participants but the study used user-led rating system in the app store.

**Primary and secondary outcome measures** BCTs and user rating.

**Results** Of 400 apps, 156 were eligible and 65 were randomly selected, downloaded and assessed by two reviewers. There was no relationship overall between star ratings and the number of BCTs present, nor between star ratings and the presence of BCTs known to be effective in increasing PA. App store was strongly associated with star ratings, with lower likelihood of finding 4 or 5 stars in iTunes compared with Google Play (OR 0.74, 95% CI 0.73 to 0.76,  $p < 0.001$ ).

**Conclusions** The findings of this study suggest that popularity does not necessarily imply the likelihood of effectiveness. Hence, public health impact is unlikely to be achieved by allowing market forces to 'prescribe' what is used by the public.

### INTRODUCTION

The accessibility, convenience and wide reach of apps create new avenues for health behaviour change on a large scale. Out of the total 325 000 health apps available on the market in 2017,<sup>1</sup> the largest app groups were fitness apps. The rapidly increased market supply of the apps reflects public demand for the new means of engaging in health

### Strengths and limitations of this study

- First study to assess the relationship between popularity of physical activity apps and their likely efficacy in two major app stores.
- A systematic approach to sample identification and assessment including a standardised assessment tool used to characterise behaviour change interventions.
- Sample identified and assessed by two independent reviewers.
- It is unknown what variables comprise the ranking algorithm from which the sample was derived.
- It is possible that user ratings, as expressed by the stars assigned to the apps, can relate to different aspects of the app content and functionalities.

behaviours. In 2015, 34% of mobile phone owners had at least one health-related app downloaded on their mobile phone.<sup>2</sup> Yet, 28.7% of adults in England are inactive.<sup>3</sup> This discrepancy may be explained by the difference between the intention to increase physical activity (PA) and the actual engagement in the behaviour, that is, the intention-behaviour gap.<sup>4</sup>

Researchers assessing the relationship between the popularity of apps and their quality have found mixed results. Azar *et al*<sup>5</sup> found that apps for weight management that were of higher quality, defined in their study as inclusion of the constructs from four behaviour change theories, were not among the highest ranked apps in the app stores. Similarly, apps that had higher download rates or higher ranking were associated with less adherence to guidelines in smoking cessation apps.<sup>6</sup> On the other hand, Pereira-Azevedo *et al*<sup>7</sup> reviewed the descriptions of 129 urology apps in the Google Play app store and found that higher download rates were associated with expert involvement in the development of the apps. These studies targeted different

BMJ

Bondaronek P, *et al.* *BMJ Open* 2019;**9**:e027536. doi:10.1136/bmjopen-2018-027536

1



health conditions and behaviours from PA, the subject of this study, and used different definitions of quality, such as consistency with behaviour change theory, expert involvement and adherence to guidelines, hence their findings need to be interpreted with caution. None of these studies used behaviour change theory to systematically assess the content of the apps in terms of their likely efficacy and how it relates to app popularity.

The Behaviour Change Technique Taxonomy (BCT-Taxonomy)<sup>8</sup> is a systematic method used to specify the content of behaviour change interventions.<sup>9</sup> The Taxonomy has also been used to quantify the inclusion of behaviour change theory in interventions, including apps.<sup>10–12</sup> For example, Bardus *et al*<sup>10</sup> extracted BCTs from 23 apps aimed at weight management. They used the Mobile App Rating Scale (MARS) to assess the quality of apps<sup>13</sup> and showed a positive association between the number of BCTs and the MARS subsets: engagement, functionality and aesthetics, as well as the overall quality of app score. This may suggest that the inclusion of BCTs may be related to the quality of the apps, as assessed by MARS.

However, the same authors found no association between user ratings and either BCTs or the quality indicators on the MARS suggesting that apps that were highly rated were not of high quality. Similarly, Crane *et al*<sup>11</sup> showed that only one BCT, 'prompt review of goals', was associated with user ratings in alcohol reduction apps. The authors concluded that there was little association between the mention of theory, BCT inclusion and popularity of the alcohol reduction apps. These findings indicate that further work is needed to explore whether popular apps are those that are of high quality and are likely to be effective. This has relevance for public health policy as the combination of popularity with likely efficacy would suggest that apps have a potential role in promoting public health, whereas a disconnect between popularity and likelihood of efficacy would raise questions about leaving the market to guide user choices in app downloads.

Systematic reviews and meta-analyses have shown that self-regulation techniques, such as goal setting, monitoring and feedback, are effective in increasing PA<sup>14–17</sup> behaviour and they may have cumulative effects. For example, Michie *et al*<sup>5</sup> showed that self-monitoring with other self-regulatory BCTs was more effective in increasing PA than using one of those BCTs in isolation. Self-regulation has been acknowledged as important constructs in behaviour change theories, for example, in control theory<sup>18</sup> and self-regulation theory.<sup>10</sup> Hence, the presence of these BCTs can be used as an indicator for quality of those apps and a proxy measure of their likely efficacy.

#### Aim and objectives

The aim of this study was to explore the relationship between the user ratings as a marker of popularity of apps and the inclusion of BCTs as a marker of likely efficacy in

publicly available PA apps. Specifically, the primary objective was to assess the association between user ratings and the number of BCTs included in the sample of popular PA apps. The secondary objective assessed the inclusion of BCTs shown to be effective in increasing PA behaviour, in particular BCTs related to self-regulation of behaviour.

## METHODS

### Design

The study used a random sample of popular apps to determine the association between user ratings and the presence of BCTs. Descriptive data included the cost and size of the app; the number and distribution of star ratings; and the presence of BCTs.

### Data sources and collection

The sample included popular PA apps identified from 400 top-ranked free and paid apps from the Health and Fitness category (100 iTunes free+100 iTunes paid+100 Google Play free+100 Google Play paid) available in the UK version of iTunes and Google Play app stores on 17 October 2016. Apps were included if (1) the primary behaviour targeted was PA; and (2) they had stand-alone functionality. Apps that specifically targeted children were excluded. These apps had to include phrases that suggested clear targeting of children. For example, 'Yoga for kids', 'Workout for kids', and 'Fun fitness for kids', 'Toddler activities' would have been excluded. The rationale for excluding apps aimed at children was (1) children may not have access to iTunes or Google Play accounts; (2) ratings may reflect the parental ratings rather than the children's; and (3) the determinants of PA in children differ from those for adults, with family and school-based activities having a major influence.<sup>20 21</sup> However, such apps were not included in the 400 top-ranked apps. Two reviewers (PB, Ghadah Alkhalidi) reviewed each app, assessed whether specific BCTs were present and extracted relevant data. If the apps existed in both stores, then the reviewers only downloaded and assessed apps in iTunes. The detailed process of app identification and sample description is provided in ref 22. Based on the number of ratings, the proportion and the distribution of star ratings we were able to reconstitute individual-level response data within each store.

### Data extraction

#### BCT extraction as a measure of likely efficacy

The presence or absence of 93 BCTs in the BCT-Taxonomy v1 as described by Michie *et al*<sup>8</sup> was classified using three categories: absent, appears to be present but evidence is not clear (+), present beyond doubt and evidence clear (++) . The presence of self-regulation BCTs, associated with PA intervention effectiveness,<sup>14–16 16 17</sup> was classified using the BCT-Taxonomy grouping 1: goal and planning, and grouping 2: feedback and monitoring, as these groups reflect the self-regulation BCTs in the Taxonomy.



### User ratings

We extracted the star ratings (1–5 stars) assigned to the apps in both stores and the number of ratings assigned to each app. The average star rating was calculated by summing the number of stars awarded across all users and dividing them by the number of users that submitted the rating. When an app appeared in both stores, a weighted average of the star ratings for each store was calculated based on the relative proportion of the ratings in each store. This algorithm is equivalent to summing the number of stars awarded by all users in both stores and then dividing them by the total number of reviews across both stores. This calculation weights users equally regardless of the platform used to access the app. The variable was dichotomised into high (4, 5 stars) or low (1, 2, 3 stars) rating.

Variables determined a priori potentially to affect the relationship between BCTs and higher or lower ratings were app store (iTunes or Google Play), number of features, whether the app was free or required payment, size (in megabytes) and usability.

### Number of features

Health apps use technology-enhanced features to deliver BCTs in order to influence behaviour. The apps were categorised according to the features offered by the app, for example, PA tracking, reminders, app community, data visualisation, and so on. To the authors' best knowledge, at the time of writing, there was no standardised list of features that are commonly used in PA apps. The first 10 apps extracted were used to compile a list of PA app features. The list was continuously updated throughout the app extraction process in order to accommodate for new features that were found in the PA app sample. The features were extracted by two reviewers (PB and GA, and any discrepancies were resolved by comparing the results of the extraction and reaching consensus).

### Usability

Usability was assessed using the System Usability Scale (SUS),<sup>23</sup> which consists of 10 items ranked on a 5-point Likert scale, from 'strongly disagree' to 'strongly agree'. To make the wording of the scale more applicable to the study, two changes were made: (1) the wording of the eighth statement was changed from 'cumbersome' to 'awkward' as recommended<sup>24–26</sup>; and (2) the word 'system' was replaced by 'app'. SUS yields scores from 0 to 100 where higher scores indicate greater usability. The interpretation of the SUS score used the thresholds proposed and validated by Bangor *et al.*<sup>24</sup> with 72.5 described as good usability.

### Data analysis

Agreement statistics including the prevalence and bias-adjusted kappa (PABAK)<sup>27</sup> statistic and unadjusted kappa<sup>28</sup> were calculated. Disagreements were then resolved by discussions among the two reviewers and consultation with other authors if unresolved.

The number of BCTs in the apps was summarised using the mean, SD, median, 25th and 75th percentiles, and the maximum and minimum. Similar statistics were used to summarise user ratings, cost, size and SUS score. The summary descriptive tables were presented for free and paid apps separately and in total as app stores have separate rankings based on the cost.

The prespecified primary analysis was based on a linear regression of the number of BCTs on star average as this continuous outcome required the smallest sample of apps for adequate statistical power. Logistic regression to determine the relationship between the number of BCTs and the odds of high and low star ratings was the prespecified secondary analysis. We modelled a higher (vs lower) star rating as the event.

### Sample size calculations

The sample size calculation was based on a pilot sample of 10 apps (five paid from iTunes, five free from Google Play) selected from the 400 apps identified. Apps were sorted in order of store rankings. From iTunes, 38 potentially eligible paid apps were identified and every eighth app was included in the pilot sample ( $n=38/5=7.6\approx 8$ ). From Google Play, 55 potentially eligible free apps were identified and every 11th app was included in the pilot sample ( $n=55/5=11$ ). If a sample app was downloaded and found to be ineligible, the next lowest ranked app was used instead. Three apps from Google Play and none from iTunes were found to be ineligible and replaced.

A pilot study based on the 10 apps suggested that mean star ratings were normally distributed. Hence, a sample size calculation was undertaken using mean ratings as a continuous measure in a linear model. Based on this framework, a sample size of 51 apps would provide 90% power to detect a change of 0.11 for each additional BCT at 5% significance level (type I error rate). A sample size of 65 apps was selected to account for any randomly selected apps that might not fulfil the inclusion criteria once downloaded. However, the pilot sample was not sufficiently large enough to show that there was a skew in star average for the complete sample once data were extracted, which would preclude the use of linear models. The difference in star rating by store was also unknown at the time of study planning, and hence the app store interactions were a post hoc addition to this plan. A retrospective power analysis was conducted. Based on simulations, this study had 97% power to detect an OR $\geq$ 1.2 with 64 apps. Since we had high power and did not see a significant result, this suggests that the true increase in the chance of a star rating of 4 or 5 with each additional BCT has an OR $<$ 1.2.

As indicated in the analysis plan, star ratings of 4 and 5 were classified as high ratings, while ratings of 1, 2 and 3 were classified as low ratings. We used the user ratings data from each store in regression models with a random effect to account for store differences. Modelling proceeded in a series of steps. First, we tested a small set of variables identified in the analysis plan for



the relationship to star rating (cost, usability, number of features, size) by including each of these as fixed covariates in univariate logistic regression models with the outcome of high versus low ratings. Any variables significant at  $p \leq 0.1$  were included in models examining BCTs. App store and the app store-by-BCT interaction terms were included in each model to reflect that the clientele for iTunes and Google Play may be different, and that these differences may impact the relationship between BCTs and star ratings. Weighted logistic regression (by the number of responses for each app) was used with a random intercept term for app, reflecting that the apps selected for analysis were a random sample of available PA apps, and to control for correlation when the same app was reviewed in both stores. Analyses were performed using SAS V.9.3 and R V.3.3.3.

#### Sensitivity analysis

Sensitivity analyses included a revision of the number of BCTs including only those that were present beyond any doubt (++),<sup>29</sup> and dichotomising user ratings into 5 stars versus  $< 5$  stars.

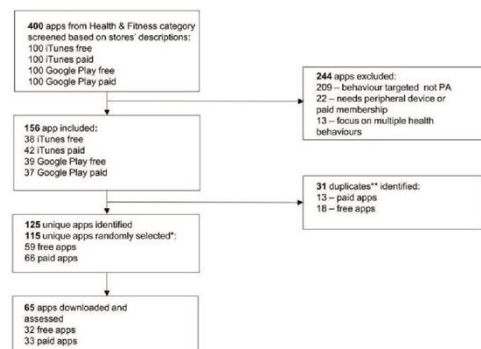
#### Patient and public involvement

This research was conducted without patient involvement.

## RESULTS

### Sample

Out of 400 apps, 244 were excluded (see figure 1). Of the remaining 156 PA apps, 31 were duplicates in that the same app appeared in both iTunes and Google Play. Thus, 125 unique apps were eligible for random selection; 65 unique apps were randomly selected for analysis including 32 free apps and 33 paid apps. One app, Break, was excluded from analyses of user reviews as the data on user ratings were not available due to the small number



**Figure 1** Flow chart of the apps included in the analysis. PA, physical activity. \* 10 apps were assessed prior to conduct sample size calculations, hence 115 were randomly selected for the sample; \*\* duplicates were defined as the same app occurring in both stores.

**Table 1** Characteristics of PA apps included in the sample

	Free (n=32)	Paid (n=32)	Total (n=64)
<b>BCTs (n)</b>			
Mean±SD	6.56±2.99	7.56±2.87	7.06±2.95
Median	7.00	8.00	8.00
25th, 75th percentiles	5.00, 8.00	6.00, 10.00	5.00, 9.00
<b>Star average*</b>			
Mean±SD	4.26±0.65	4.35±0.45	4.30±0.56
Median	4.50	4.49	4.50
25th, 75th percentiles	4.30, 4.60	4.29, 4.60	4.30, 4.60
<b>Features (n)</b>			
Mean±SD	5.94±2.17	5.72±1.95	5.8±2.05
Median	7.00	8.00	6.00
25th, 75th percentiles	4.00, 8.00	5.00, 8.00	4.00, 8.00
<b>Size (MB)*</b>			
Mean±SD	45.53±40.85	61.51±47.22	53.52±44.53
Median	32.45	53.97	39.64
25th, 75th percentiles	16.82, 62.27	28.92, 85.02	25.52, 79.50
<b>Usability (SUS)</b>			
Mean±SD	81.25±12.64	85.23±12.02	83.24±12.40
Median	85.00	87.50	86.25
25th, 75th percentiles	71.88, 91.25	78.75, 93.75	75.00, 91.88
<b>Cost (£)*</b>			
Mean±SD		3.00±2.12	
Median	N/A	2.35	N/A
25th, 75th percentiles		1.78, 2.99	

\*weighted averages of iTunes and Google Play. BCT, Behaviour Change Techniques; MB, megabyte; N/A, not applicable; PA, physical activity; SUS, System Usability Scale.

of ratings, resulting in a final sample of 64 PA apps. These apps collectively received more than 2.8 million user ratings. Individual-level data of the characteristics of each app are provided in Bondaronek *et al's*<sup>32</sup> study.

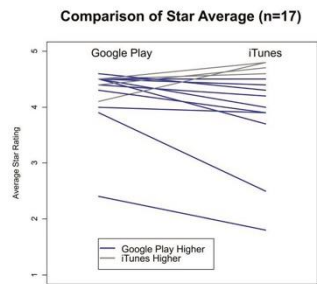
There was substantial agreement between the reviewers in the assessment of BCTs (PABAK=0.94, 95% CI 0.93 to 0.95, kappa=0.78, 95% CI 0.75 to 0.81).

Characteristics of the apps are displayed in table 1. The number of BCTs and features was approximately normally distributed, with the mean number of BCTs being 7.1 (SD=3.0) and the mean number of features being 5.8 (SD=2.1) (see online supplementary file 1 for list of PA app features and frequency of occurrence). Star ratings, size, SUS score and the cost variables were skewed. The median star rating was 4.5 (IQR 1.9–4.9) and the median SUS score was 86.3 (IQR 75.00–91.88). Among paid apps, the median cost was £2.40 (IQR 1.78–2.99).

### User ratings

In total, 2 819 469 user ratings of the 64 apps were used in the analysis. Among these, 88.5% were 4 or 5-star reviews and 11.5% were 1, 2 or 3-star reviews. Among the covariates considered for model inclusion, good usability





**Figure 2** The comparison of star average between the same apps that existed in both stores (n=17).

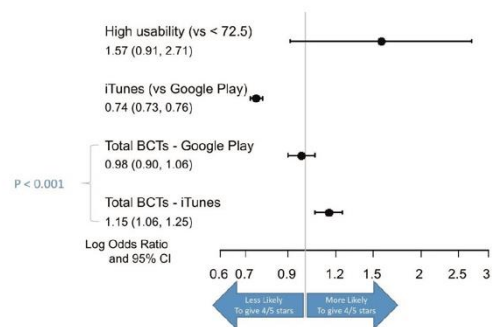
increased the chance of a 4 or 5-star rating compared with poor usability (OR 1.66, 95% CI 0.96 to 2.89,  $p=0.071$ ). This covariate was included in all subsequent models. App store was strongly associated with star ratings, with lower likelihood of finding 4 or 5 stars in iTunes compared with Google Play (OR 0.74, 95% CI 0.73 to 0.76,  $p<0.001$ ).

Further investigation of the difference between the stores and user ratings showed that there was 37% higher likelihood of awarding 4 or 5 stars in Google Play compared with iTunes (95% CI 35% to 41%,  $p<0.001$ ). A similar pattern can be seen comparing the weighted averages within app star ratings across stores. There were 17 apps that were sold in both stores (figure 2). Analysis was conducted to assess whether the proportion with the high star ratings in Google Play exceeds what would have been found by chance. If there were no relationship between store and star rating, we would expect the star rating in iTunes to exceed the star rating in Google Play for about 50% of apps (coin flip). Mean star rating in Google Play exceeded the rating in iTunes in 13/17 apps (76.5%, 95% CI 64.2% to 86.2%,  $p<0.0001$ ). Since the analysis was significant, we can rule out that the difference between the user ratings in the stores was due to chance. Cost, size and number of features were not related to star ratings.

**Primary analysis**

The primary analysis of the association between user ratings and the overall number of BCTs included in each app found no relationship between number of BCTs and star rating (OR 1.05, 95% CI 0.97 to 1.14,  $p=0.236$ ). Subgroup analysis showed that in iTunes only, a higher star rating was associated with the total number of BCTs (OR 0.72, 95% CI 0.71 to 0.74,  $p<0.001$ ).

Based on the model containing the store interaction (figure 3), in iTunes there was an association with each additional BCT corresponding to 15% increase in the likelihood of a higher star rating (OR 1.15, 95% CI 1.06 to 1.25), there was no association between the number of BCTs and Google Play. Usability was not significant in the multivariate model (OR 1.57, 95% CI 0.91 to 2.71).



**Figure 3** ORs for the associations between the number of BCTs and a 4 or 5-star rating. BCT, Behaviour Change Techniques.

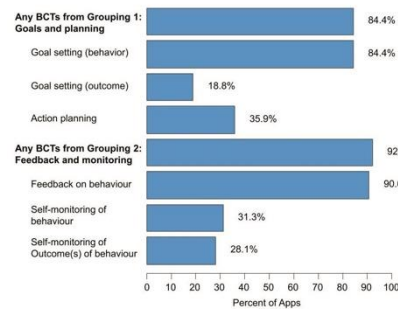
**Secondary objective**

The secondary analysis examined the associations between a 4 and 5-star user rating and the inclusion of BCTs that have been shown to be effective in increasing PA behaviour.

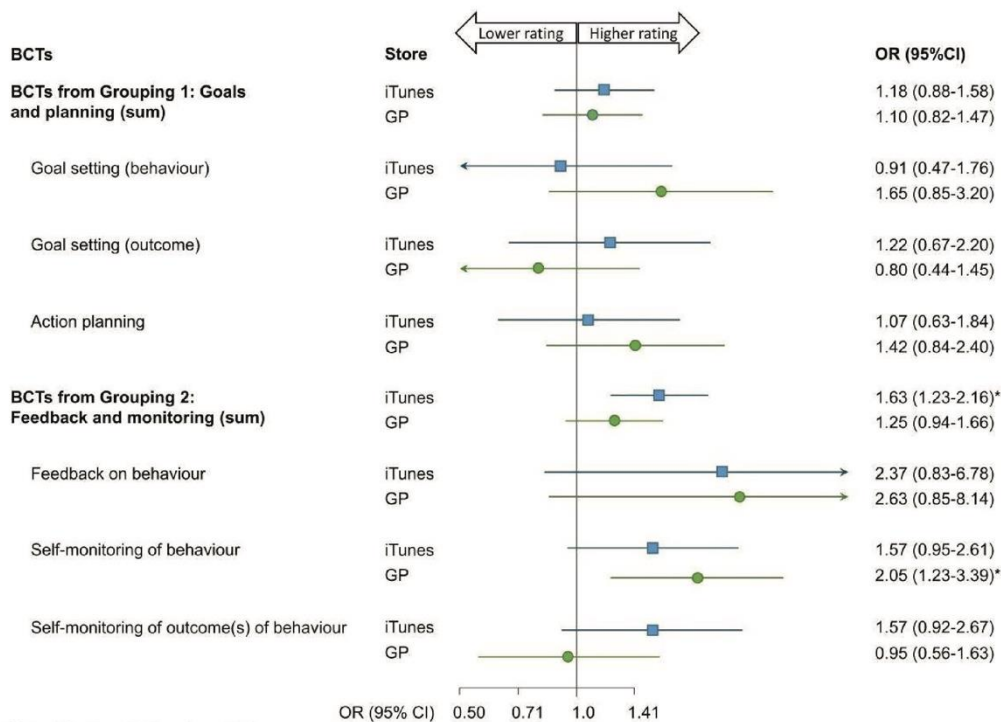
The frequency of the BCTs associated with efficacy for increasing PA behaviour is presented in figure 4.

The BCT grouping 1 (goal and planning) was represented in 84.4% of evaluated apps. The most common of these was goal setting for behaviours (84.4%), followed by action planning (35.9%) and goal setting for outcomes (18.8%). The BCT grouping 2 (feedback and monitoring) was included in 92.2% of evaluated apps. Feedback on behaviour was included in nearly all apps (90.6%), while self-monitoring of behaviour and outcomes was incorporated in about a third of apps.

The number of grouping 1 BCTs had no impact on star rating in either store (figure 5). Among iTunes users, each additional BCT from grouping 2 increased the chance of a high star rating by 63% (OR 1.63, 95% CI 1.23 to 2.16). There was no association between the number of BCTs and star rating among Google Play users. Self-monitoring



**Figure 4** Frequency of individual BCTs within the two groupings of self-regulatory BCTs. BCT, Behaviour Change Techniques.



**Figure 5** ORs for the associations between BCTs that have been shown to be effective and a 4 or 5-star rating.  $p < 0.05$ . BCT, Behaviour Change Techniques; GP, Google Play.

of behaviour was associated with higher ratings in Google Play (OR 2.05, 95% CI 1.23 to 3.39).

#### Sensitivity analysis

The sensitivity analyses including only those BCTs that were classified as 'present beyond all reasonable doubt'<sup>29</sup> showed that for iTunes, the BCT 2.2 feedback on behaviour crossed into significance. Otherwise, the results are consistent with the original findings. The results of the sensitivity analysis are available in the online supplementary file 2.

## DISCUSSION

### Statement of principal findings

The aim of this study was to assess the relationship between the popularity of publicly available PA apps (assessed through user ratings) and their likely efficacy (assessed through the inclusion of BCTs). Overall, for both app stores, there was no association between popularity and likely efficacy as indicated by the overall presence of BCTs and the BCTs known to be effective in increasing PA. However, users in each app store differed and there

was an association between the number of BCTs and high user rating in iTunes but not in Google Play.

### Strengths and weaknesses of this study

The main strength of the study includes the systematic assessment of the app content conducted by two reviewers. The sample was identified from the most popular publicly available apps from two major app distribution platforms. The use of BCT-Taxonomy provides a standardised assessment tool, which has been used in other studies assessing the content of apps.<sup>12, 30-32</sup> Second, Guzman<sup>33</sup> argued that the star rating represents an average score for the whole app that combines both positive and negative evaluations aggregated across users. The study, however, used rater-level data which included individual ratings from 2.8 million users. These large numbers mitigate the problem posed by averaging the star ratings across users.

This study has some important limitations. First, the main limitation of the study relates to the variables used in this study. It is possible that user ratings, as expressed by the stars assigned to the apps, can relate to different aspects of the app functioning and content. There is evidence suggesting that app reviews tend to occur near a



new release which can suggest that the ratings may include comments on the specific updates of the software.<sup>34</sup> In addition, the possibility that user ratings were influenced by fake reviews cannot be excluded.<sup>35, 36</sup> However, the user rating was considered the most appropriate measure to explore since it represents a user-led feedback that reflects user experience. Similarly, the choice of the BCTs as an approximation for likely efficacy was selected because studies assessing the efficacy of the apps on the market are scarce.<sup>37–39</sup> Second, the ranking algorithm from which the sample was derived is unknown. Hence, this lack of transparency prevented evaluation of how the calculation of rank might have influenced the app selection. However, apps appear in rank order by default in app stores, hence the rank affects what users are seeing. As the aim of the study was to assess the most popular apps, the choice of highly ranked apps was considered the most appropriate for the context of this research. Third, Google Play market tends to have more ratings than iTunes because the process of app review is more complex in the latter.<sup>40</sup> This was addressed in the study by using the weighted averages of the ratings across the stores (for the summary of the app characteristics), and by controlling for store in the regression models. While there is a difference in the app review process in both app stores, feedback from both stores should be recognised as valid and important. Fourth, while we inspected the apps to ensure that the duplicate apps were similar in both stores, there could be differences in the functionality of the apps between the app stores that we could not see. Fifth, failure to detect the skew in the original primary outcome used to power the study is a limitation. However, a retrospective power analysis showed we had high power to detect an OR of 1.2 and did not find significant result, hence the true OR is likely to be <1.2. Lastly, the sample identification was obtained in October 2016 reflecting an extract of the state of the market at the time.

#### Strengths and weaknesses in relation to other studies, discussing particularly any differences in results

This study supports the findings of previous research in apps targeting weight management,<sup>5, 10</sup> smoking cessation<sup>6</sup> and alcohol reduction<sup>11</sup> which showed that apps that were highly rated, highly ranked or highly downloaded were not necessarily of high quality. In addition, the inclusion of BCTs that have been shown to be effective in increasing PA, that is, the self-regulation strategies, was also not associated with higher ratings and this result supports the similar findings of Bardus *et al.*<sup>10</sup> No association between the behaviour change theory content and user ratings suggests other factors may be contributing to the apps' popularity. High-quality graphic design, visual appeal and ease of use are more likely to attract potential customers to download and engage with the app.<sup>41</sup> In addition, the promotion of the apps can play a role in the download rates. The strength of this study, in comparison to other handful research exploring this association, is the use of logistic regressions, including the analysis

of the potential confounding factors, provides a strong evidence for the association, if such was existing.

The finding of an association between user rankings and overall number of BCTs in one store (iTunes) but not the other (Google Play) is new and one that has not been found in previous studies. There are various possible explanations for this finding. Market researchers suggest possible differences in the population of users of each store.<sup>42, 43</sup> For example, iPhone users might be more affluent, they engage with their device longer, make more purchases with their phone and are more loyal to their brand.<sup>44</sup> In addition, there are differences in the review and approval processes between the two stores which could have influenced the user ratings.<sup>45</sup> Future studies should explore this observation further, and determine whether it holds for health apps targeting different behaviours.

#### Meaning of the study: possible mechanisms and implications for clinicians or policymakers

In this study, we showed no evidence of association between popularity and likely efficacy. The implication of this study is that the popularity of these apps is not a sufficient filter to distinguish the apps that might have higher potential to have an impact on the potential users. Hence, we suggest that, at present, allowing the commercial market to determine which PA apps are downloaded is unlikely to be an effective method of public health promotion in terms of increasing the overall levels of PA.

Based on the findings, we suggest some implications for public health policy. Apps aimed to increase PA represent the largest category in the two major app stores<sup>46</sup> which illustrates public demand for engaging in PA. The lack of quality in those apps indicates a missed opportunity to increase health at the population level. Initiatives to identify and promote high-quality apps are in development, for example, the NHS Apps Library.<sup>47</sup> However, there is an urgent need to evaluate the effectiveness of these apps with potential users.

#### Unanswered questions and future research

Further research is needed to understand which BCTs, and in which combination, are most effective in increasing PA when delivered via an app; how these BCTs can best be delivered; and how to combine features which promote efficacy with those that promote popularity. Last, the differences in iTunes and Google Play users are an unexpected finding, and not one that the study set out to identify (ie, not an a priori hypothesis). Future researchers should be aware of the potential differences between iTunes and Google Play users and ensure research is carried out on both platforms.

#### CONCLUSION

To date, this is the first study to assess the association between popularity (measured using user ratings of the apps) and likely efficacy (measured using the inclusion of the BCTs) of publicly available highly ranked PA apps available in the major app stores. No relationship was



found between popularity and likely efficacy suggesting that popularity does not assure high quality, and what is liked may not be what is likely to be effective. However, PA apps in this study were highly rated, highly ranked apps from the major app stores, hence highly visible for the potential user. Hence, promotion of public health is unlikely to be achieved by allowing market forces to determine which PA apps are used. More studies are needed to assess the effectiveness of apps with users in a real-world setting to investigate the app components that are both effective and valued by the users.

**Twitter** Paulina Bondaronek @paulinabond1

**Acknowledgements** We acknowledge Ghadah Alkhalidi (Community Health Sciences Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia) for her help with the data extraction.

**Contributors** PB designed the study, collected the data, analysed the data and wrote the manuscript. AS analysed the data and cowrote the manuscript. EM and FLH provided advice on study design, data analysis and interpretation, and contributed to the critical appraisal of the manuscript.

**Funding** PB is a PhD student funded by the Medical Research Council.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Paulina Bondaronek <http://orcid.org/0000-0003-0096-1234>  
Elizabeth Murray <http://orcid.org/0000-0002-8932-3695>

#### REFERENCES

- research2guidance. mHealth APP economics: current Status and future trends in mobile health: research2guidance, 2017. Available: <https://www.uzelf.org/wp-content/uploads/2017/12/R2G-mHealth-Developer-Economics-2017-Status-And-Trends.pdf> [Accessed 20 Sep 2018].
- Terry K. Number of health apps soars, but use does not always follow, 2015. Available: <http://www.medscape.com/viewarticle/851226>
- Public Health England. Public health outcomes framework: physical activity, 2016. Available: <http://www.phoutcomes.info/search/physical%20activity#pat/6/ati/102/par/E12000004>
- Rhodes RE, Dickau L. Moderators of the intention-behaviour relationship in the physical activity domain: a systematic review. *Br J Sports Med* 2013;47:215–25.
- Azar KMJ, Lesser LI, Laing BY, et al. Mobile applications for weight management: theory-based content analysis. *Am J Prev Med* 2013;45:583–9.
- Abroms LC, Padmanabhan N, Thaweethai L, et al. iPhone apps for smoking cessation: a content analysis. *Am J Prev Med* 2011;40:279–85.
- Pereira-Azevedo N, Osório L, Cavadas V, et al. Expert involvement predicts mHealth APP downloads: multivariate regression analysis of urology apps. *JMIR Mhealth Uhealth* 2016;4.
- Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (V1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013;46:81–95.
- Michie S, Atkins L, West R. The behaviour change wheel. In: *A guide to designing interventions*. 1st edn. Great Britain: Silverback Publishing, 2014.
- Bardus M, van Beurden SB, Smith JR, et al. A review and content analysis of engagement, functionality, Aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *Int J Behav Nutr Phys Act* 2016;13.
- Crane D, Garnett C, Brown J, et al. Behavior change techniques in popular alcohol reduction apps: content analysis. *J Med Internet Res* 2015;17:e118.
- Middelweerd A, Mollee JS, van der Wal CN, et al. Apps to promote physical activity among adults: a review and content analysis. *Int J Behav Nutr Phys Act* 2014;11.
- Stoyanov SR, Hides L, Kavanagh DJ, et al. Mobile APP rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015;3:e27.
- Greaves CJ, Sheppard KE, Abraham C, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. *BMC Public Health* 2011;11:119.
- Michie S, Abraham C, Whittington C, et al. Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychol* 2009;28:690–701.
- Conn VS, Hafidahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. *Am J Public Health* 2011;101:751–8.
- Pearson ES. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. *Patient Educ Couns* 2012;87:32–42.
- Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. *Psychol Bull* 1982;92:111–35.
- Kanfer FH. Implications of a self-regulation model of therapy for treatment of addictive behaviors. *treating addictive behaviors*. Springer, 1986: 29–47.
- Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. *Med Sci Sports Exerc* 2000;32:963–75.
- Hesketh KR, O'Malley C, Paes VM, et al. Determinants of change in physical activity in children 0–6 years of age: a systematic review of quantitative literature. *Sports Medicine* 2017;47:1349–74.
- Bondaronek P, Alkhalidi G, Slee A, et al. Quality of publicly available physical activity Apps: review and content analysis. *JMIR Mhealth Uhealth* 2018;6.
- Brooke J. SUS-A quick and dirty usability scale. usability evaluation in industry 1996;189:4–7.
- Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum Comput Interact* 2008;24:574–94.
- Finstad K. The system usability scale and non-native English speakers. *J Usability Stud* 2006;1:185–8.
- Lewis J. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int J Hum Comput Interact* 2002;14:463–88.
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–9.
- Cohen J. Kappa: coefficient of concordance. *Educ Psych Measurement* 1960;20.
- Wood CE, Richardson M, Johnston M, et al. Applying the behaviour change technique (BCT) taxonomy V1: a study of coder training. *Transl Behav Med* 2015;5:134–48.
- Direito A, Pfaeffli Dale L, Shields E, et al. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? *BMC Public Health* 2014;14:1.
- Chen J, Cade JE, Allman-Farrell M. The most popular smartphone Apps for weight loss: a quality assessment. *JMIR Mhealth Uhealth* 2015;3:e104.
- Yang C-H, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. *Am J Prev Med* 2015;48:452–5.
- Guzman E, Maalej W. *How do users like this feature? A fine grained sentiment analysis of APP reviews*. Karlskrona, Sweden: IEEE, 2014.
- Pagano D, Maalej W. User feedback in the appstore: an empirical study. requirements engineering conference (re), 2013 21st IEEE international 2013.
- Girardello A, Michahelles F. AppAware: which mobile applications are hot? proceedings of the 12th International Conference on human computer interaction with mobile devices and services 2010.
- Frost VS, Kuehnhausen M. Trusting smartphone apps? to install or not to install, that is the question. cognitive methods in situation awareness and decision support (CogSIMA) 2013.
- Direito A, Carraça E, Rawstorn J, et al. mHealth technologies to influence physical activity and sedentary behaviors: behavior change





- techniques, systematic review and meta-analysis of randomized controlled trials. *Ann Behav Med* 2017;51:226–39.
- 38 Buechi R, Faes L, Bachmann LM, et al. Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis. *BMJ Open* 2017;7:e018280.
- 39 Jake-Schoffman DE, Silfee VJ, Waring ME, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. *JMIR Mhealth Uhealth* 2017;5.
- 40 Xu W, Liu Y. mHealthApps: a Repository and database of mobile health Apps. *JMIR Mhealth Uhealth* 2015;3:e28.
- 41 Lazard AJ, Mackert MS. E-Health first impressions and visual evaluations: key design principles for attention and appeal. *Communication Design Quarterly Review* 2015;3:25–34.
- 42 Götz FM, Stieger S, Reips U-D. Users of the main smartphone operating systems (iOS, android) differ only little in personality. *PLoS One* 2017;12:e0176921.
- 43 Monetate. Ecommerce Quarterly 2016: mobile ate the world and is still hungry, 2017. Available: [http://info.monetate.com/rs/092-TQN-434/images/Monetate-EQ3-2016.pdf?mkt\\_tok=eyJpIjoiTURFMFotTTNNR1prWWpRMylsInQiOiJsaDE4bXF1OTA4R0c4T0xjNGZlXC9JRkxcHM4WkRTZFZkT2tnWWorMGtjc24zSUJfsanN2Z2tOVDJHtDZFc3F1eTZNZEZJaWJsQUxzQWRjN3F6ZXVnc2wrcVREMDFJc3RMQU5yc0NGdXU4OG41Zk9Gdk1OaGZRUXNEcisrWFEczcifQ%3D%3D](http://info.monetate.com/rs/092-TQN-434/images/Monetate-EQ3-2016.pdf?mkt_tok=eyJpIjoiTURFMFotTTNNR1prWWpRMylsInQiOiJsaDE4bXF1OTA4R0c4T0xjNGZlXC9JRkxcHM4WkRTZFZkT2tnWWorMGtjc24zSUJfsanN2Z2tOVDJHtDZFc3F1eTZNZEZJaWJsQUxzQWRjN3F6ZXVnc2wrcVREMDFJc3RMQU5yc0NGdXU4OG41Zk9Gdk1OaGZRUXNEcisrWFEczcifQ%3D%3D) [Accessed 21 Sep 2018].
- 44 Katariya J. Apple vs android — a comparative study, AndroidPub, 2017. Available: <https://android.jelise.eu/apple-vs-android-a-comparative-study-2017-c5799a0a1683>
- 45 Petersen K, Gonzalez-Huerta J, Ickin S. Why do users install and delete Apps? A survey study. *International Conference of Software Business* 2017;304:186–91.
- 46 research2guidance. Mhealth APP developer economics. The state of the art of mHealth application publishing: research2guidance, 2014. Available: <http://research2guidance.com/product/mhealth-app-developer-economics-2014/> [Accessed 5 May 2016].
- 47 NHS. Nhs Apps library. Available: <https://www.nhs.uk/apps-library/> [Accessed 16 Jun 2019].