University of Southampton

Faculty of Environmental and Life Sciences

School of Psychology

Understanding digital intervention engagement: making sense of large-scale data

By

Sascha Jane Miller

ORCID ID 0000-0002-1949-5774

Thesis for the degree of Doctor of Philosophy in Psychology

July 2021

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

School of Psychology

Thesis for the degree of Doctor of Philosophy

Understanding digital intervention engagement: making sense of large-scale data

by

Sascha Jane Miller

Digital behaviour change interventions provide convenient and personalised health support for users, and the opportunity to record substantial amounts of data about users' interactions with the intervention. If analysed systematically, these data are able to explain how the intervention was effective, for whom and in what context, leading to recommendations on how the intervention and future dissemination may be improved. However, the volume of data, their complexity and diversity can become a barrier. The aim of this thesis was to devise and apply a method to support analyses of large-scale usage data.

The framework for Analysing and Measuring Usage and Engagement Data (AMUsED) was developed to support researchers in establishing clear rationales for collecting usage metrics and undertaking inferential analyses. The framework was applied to usage analyses of two interventions addressing antimicrobial resistance by lowering unnecessary medication prescriptions: Internet Dr encourages self-care for respiratory tract infections thereby reducing avoidable GP visits; PRIMIT/Germ Defence intervention lowers transmission of viruses by increasing handwashing in the home. The process evaluations identified: what type of engagement was successful; specific improvements for the interventions; and the importance of context for using the intervention. Internet Dr findings revealed that the intervention is effective at raising enablement to self-care for users who are not experiencing symptoms, suggesting the structure and theory-based content are relevant for increasing self-care for other minor ailments. The PRIMIT/Germ Defence findings provide insight for public health campaigns by evidencing the value of targeting multiple handwashing situations where risk of transmission is high.

A scoping review was carried out to capture how 'usage' and 'engagement' with digital interventions are conceptualised. The review confirms the need for a practical and generalisable method for usage analyses, and the AMUsED framework is the only method proposed to do this.

The process evaluations structured by the framework demonstrate how it supports researchers in conducting analyses of large-scale data, leading to a better understanding of digital intervention engagement, specific recommendations for improvement, and informs our understanding of applying behaviour change theories to a target health issue. These types of findings will lead to more effective digital behaviour change interventions that provide better support for the people who use them.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Sascha Jane Miller

Title of thesis: Understanding digital intervention engagement: making sense of large-scale data

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly while in candidature for a research degree at this University;
2. No part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

   Miller, S., Ainsworth, B., Yardley, L., Milton, A., Weal, M., Smith, P., & Morrison, L. (2019). A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. *Journal of Medical Internet Research*, *21*(2), e10966. https://doi.org/10.2196/10966

   Miller, S., Ainsworth, B., Weal, M., Smith, P., Little, P., Yardley, L., & Morrison, L. (2021). Process evaluations of a web-based intervention to increase handwashing during a pandemic: Moving Germ Defence from a randomised controlled trial to public dissemination. *Journal of Medical Internet Research,* 26104 (forthcoming/in press).

Signature: Date: 30th July 2021

Acknowledgements

I would like to thank the Economic and Social Research Council for funding this research (award: 1692136).

I would like to thank all my co-authors for their expertise and support, and particularly my supervisors Peter Smith and Mark Weal for their encouragement and knowledge, and for challenging and guiding me. I would also like to thank Lucy Yardley for her support over the years and the amazing opportunity of being part of the Germ Defence team. I am always inspired by her ability and commitment to deliver interventions that people want to use. Finally, I am incredibly grateful to Leanne Morrison for her help over the last five years. Coming back to a PhD was very daunting, but she has always been encouraging, insightful, inspiring and patient.

My love of psychology comes from Dad who always saw the best in people.

Abbreviations

AMUsED	Analysing and Measuring Usage and Engagement Data
AMR	Antimicrobial Resistance
BCT	Behaviour Change Technique
CONSORT-EHEALTH	Consolidated standards of reporting trials of electronic and mobile health applications and online teleHealth
DBCI	Digital Behaviour Change Intervention
GP	General Practitioner
HCP	Healthcare Professional
JITAI	Just-In-Time Adaptive Interventions
LVT	LifeGuide Visualization Tool
MOST	Multiphase Optimization Strategy Trial
MRC	Medical Research Council
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PBA	Person-Based Approach
PBC	Perceived Behavioural Control.
PEI	Patient Enablement Index
PETS	Problematic Experiences of Therapy Scale
PRIMIT	Primary Care Trial of a Website-based Infection Control Intervention to Modify Influenza-like Illness and Respiratory Infection Transmission
RCT	Randomised Controlled Trial
RTI	Respiratory Tract Infection

SMART	Sequential Multiple Assignment Randomised Trial
TIDieR	Template for Intervention Description and Replication
TPB	Theory of Planned Behaviour

Chapter 1 Research Introduction

1.1 Research aim and objectives

This research was funded by the Economic and Social Research Council as part of a commissioned call to use pre-existing secondary data sets. The primary aim of this thesis was to devise a methodology that can guide and enhance the analysis of large-scale data collected from digital behaviour change interventions (DBCIs). Applying this methodology was intended to lead to usage analyses that provide findings with insight into engagement with the interventions. The following objectives were identified and addressed:

- To explain how large-scale data from digital health interventions can be best utilised to develop our understanding of the role of engagement in positive outcomes.
- To develop a method for analysing large-scale data that is systematic and efficient, and produces findings that are meaningful to the intervention and inform our understanding of how theoretically-underpinned interventions are used.
- To enable the consideration of appropriate data analyses and required data collection during the development stage of a digital intervention, and to support post-hoc analyses of pre-collected data.
- To apply and test the methodology through DBCI process evaluations demonstrating how insights may be gained for understanding intervention efficacy, leading to suggestions for improvement, and informing the wider field of engagement with digital interventions.
- To describe concepts currently used to define and analyse usage, and critically discuss how they have been applied to DCBIs, to demonstrate the contribution of the PhD research within a rapidly developing field.

1.2 Background

1.2.1 The role of process evaluations of DBCIs

The advancement of digital technology over the last three decades has seen a proliferation of DBCIs developed, trialled and disseminated, with the aim of supporting positive health behaviour change and management of health conditions (Arigo et al., 2019). Digital interventions can empower users to self-manage their health and overcome challenges to accessing support (e.g. time, privacy), whilst reducing costs for providers (Arigo et al., 2019). Primary evaluation of DBCIs typically focuses on whether intended outcomes have been achieved. However, as DBCIs have become increasingly complex, in-depth process evaluations are encouraged to examine what parts of an intervention work, for whom and under what conditions (Craig et al., 2008; Eysenbach & CONSORT-EHEALTH Group, 2011; Moore et al., 2014; Moore et al., 2015). Understanding these aspects is crucial for advancing both the specific intervention and the wider areas of DBCI development and engagement.

Recommendations for process evaluations propose using mixed methods combining qualitative data capturing users' subjective experiences (e.g. acceptability, usability, reactions to the intervention and offline behaviour; Michie et al., 2017; Yardley et al., 2016), and quantitative data accrued via interaction with the intervention (e.g. intervention usage, self-report data, user characteristics, behavioural outcomes). When applied in an integrated manner these become additive, with quantitative data testing hypotheses derived from qualitative findings, and qualitative data explaining phenomenon observed in quantitative data (Moore et al., 2015). Whilst the application of qualitative methods in process evaluations faces its own challenges and debates (Cheng & Metcalfe, 2018), the focus of this research is on the potential contribution of quantitative data, and specifically usage analyses.

Analyses of intervention usage data are an essential part of an in-depth process evaluation. Usage data provides an opportunity that is unique to DBCIs: to examine objective, anonymous recorded interactions of individual users with the intervention (e.g. content viewed, time spent, number of logins) (Danaher et al., 2006). Whilst constructing digital interventions it is possible to embed data collection into the intervention structure to automatically capture a wealth of information, whilst requiring minimal further effort from the participants and the research team (Danaher et al., 2006; Yardley et al., 2016). If analysed meaningfully, the findings can lead to improved intervention architecture and implementation methods, with the potential to increase positive outcomes, lower burden on users, and reach specific groups more successfully (Danaher, Boles, Akers, Gordon, & Severson, 2006; Eysenbach, 2005; Michie, Yardley, West, Patrick, & Greaves, 2017; Yardley et al., 2016). Usage data can also improve our broader understanding of the effects of theory-based content, patterns of engagement and impact of changes in context on intervention efficacy, offering vital lessons for future intervention design and development. In addition, findings from DBCI usage have the potential to advance theories of human behaviour and inform all behavioural interventions (Michie et al., 2017).

1.2.2 Challenges to defining and measuring usage of DBCIs

Although the potential value of analysing usage data has been widely recognised for over two decades, this has not yet been realised and usage analyses are still underutilised (see Scoping Review, Chapter 2). Therefore, it is important to identify and address the barriers that stand in the way to attaining the insight they are able to provide.

Perhaps the most significant barrier is the volume of data. With many variables available, usage can be operationalised in different ways, for example: the number of times a user logs in; time spent on the website; number of pages viewed; modules or features used; completing the website. However, it is not always apparent what is being measured and what can be inferred from the interaction (Sieverink, Kelders, Poel, & van Gemert-Pijnen, 2017). For example, the amount of time a user spends on a page appears to be a clear, specific metric, yet it may indicate that the user has: high levels of engagement with the content; slow processing speed or low literacy; left their device with the webpage open but are actually elsewhere. Overall levels of engagement are equally ambiguous when trying to establish whether users have disengaged for positive or negative reasons. For example, if a user no longer visits an intervention they may have: lost interest; feel lacking in support; met their goals; feel successful (Eysenbach, 2005; O'Brien & Toms, 2008).

Research focused on usage analyses has sought to bring meaning to usage metrics by conceptualising broader ideas around intervention interaction. The Law of Attrition was one of the first models to focus on what disengagement from DBCIs and their trials can tell us about both the users and the effectiveness of the intervention (Eysenbach, 2005). However, the methods and findings in this paper were based on the assumption that disengagement is due to users' negative responses to the intervention. This premise became wide-spread in usage analyses, and descriptive statistics of usage metrics were taken as indicative of users' levels of engagement with the intervention, assuming that more usage equates to more engagement, and therefore better outcomes (Couper et al., 2010; Danaher et al., 2006).

This definition has persisted perhaps because it is analogous to the dose-response relationship seen in medicine. However, the use of inferential analyses to assess associations between usage metrics and outcomes demonstrates that the amount of content viewed is not always associated with improved outcomes. Alternatively, other patterns of usage such as the type of content viewed, the quality of the interaction, or the context in which it was accessed may be associated with behaviour change (Ainsworth et al., 2016; Donkin et al., 2013; Yardley et al., 2016). The concept of 'effective engagement' defines this as the sufficient level of engagement needed to achieve intended outcomes (Yardley et al., 2016). When establishing what type of usage constitutes effective engagement, the usage metric will necessarily need to be specific to the architecture of the intervention and the target behaviour, and may vary across contexts and user groups (Yardley et al., 2016). This means that in addition to inferential analyses examining usage and outcomes, other accrued data should also be considered such as user characteristics (Sieverink, Kelders, Poel, et al., 2017). However, with the large amount of usage metrics available, carrying out inferential analyses on all of them in order to establish effective engagement would be inefficient and difficult to interpret (Danaher et al., 2006; Michie et al., 2017).

1.2.3 Identifying and applying meaningful usage metrics

Having an in-depth understanding of an intervention makes the process of identifying appropriate usage metrics from potentially large and unwieldy data sets easier and also avoids the risk of data-dredging. Methodologically developed digital interventions draw from theory and evidence to shape the architecture, structure and content. For example, interventions developed using the person-based approach construct ‘guiding principles’ which draw together health-associated issues and relevant theory to create content that addresses these issues (Yardley, Morrison, Bradbury, & Muller, 2015). Therefore, it is anticipated that people who use these sections of intervention content will experience positive changes in those issues, providing a clear rationale for creating usage metrics that capture interaction with this specific content. The advantage of selecting measures of usage that are theory- or evidence-based is that this allows researchers to discount many variables that would not provide insights into intervention engagement (Sieverink, Kelders, & Gemert-Pijnen, 2017), thus reducing the number of variables of interest. It should be noted that this process does not discount the notion that more usage leads to better outcomes. It is apparent that for some interventions and types of health behaviours increased usage is related to improvement (e.g. the amount of content accessed or the time span of usage; Donkin et al., 2011; Kelders, Kok, Ossebaard, & Van Gemert-Pijnen, 2012). However, examining the type of content and interactions that underlie increased usage will lead to a more detailed explanation of how the intervention provides better support.

Identifying meaningful usage metrics provides several benefits when carrying out inferential analyses. Firstly, researchers are able to examine not just how an intervention was effective, but also variations amongst groups of users and in changes to the target behaviour/condition (Donkin et al., 2011; Perski, Blandford, West, & Michie, 2017; Sieverink, Kelders, & Gemert-Pijnen, 2017). Applying meaningful usage metrics also supports the process of testing specific hypotheses (Moore et al., 2015) and is more likely to discern patterns of usage that constitute effective engagement. As the findings from these types of analyses relate closely to the structure of the intervention, they are also able to provide guidance for improvements in design and future dissemination (see Chapters 4 and 5 for examples). In addition, reducing the number of usage metrics makes the process of inferential analyses less arduous and circumvents the risk of data-dredging.

1.2.4 The need for a methodology to guide usage analyses

The Scoping Review presented in Chapter 2 highlights differing viewpoints on appropriate usage metrics when reporting usage analyses within the field. This disagreement is probably the result of the lack of an existing, viable method that provides a rationale to drive selection of usage metrics. Certainly the need for one has been a shared experience, as in 2017 (whilst the current research was addressing the same issue) Sieverink et al. produced a protocol explaining: what log-data is; the value of research questions to explore usage, outcomes, context, and underpinning theory; the importance of planning data collection, preparation and analysis. This seminal paper was the first to break down these processes and provide suggestions on how to address them. However, whilst the protocol makes important distinctions using worked examples on the necessary steps involved in usage analyses, it does not contain clear guidance on how to apply these to other interventions in a systematic manner. For example, the paper emphasises the importance of forming research questions to evaluate intervention usage, yet provides no guidance on which metrics to select and how to apply them. The paper includes a limited number of example research questions that are specific to the research team's interventions, making them harder to apply to other interventions. Crucially, the research questions are based on the assumption that more usage leads to better outcomes, missing the opportunity to examine whether certain types or patterns of usage are more effective than others.

The reality of carrying out research is that the process of developing, trialling and analysing an intervention takes several years and involves many different people. Interventions are now more frequently being developed with external commercial partners who may not be familiar with the rigours of scientific analysis and the importance of usage data for providing insight into psychological theory. Anecdotal examples exist of intervention trials that have been carried out where the partner company has failed to capture anything but the most basic log-data – making in-depth usage analyses impossible. At the other end of the scale, interventions that capture large data sets in trials are time-consuming and unwieldy to analyse in detail, particularly if the analysis is to be undertaken by someone unfamiliar with the intervention. Finding the resources to trawl through these data sets can be challenging, and as a consequence they are sometimes left, missing the opportunity to gain the valuable insights held within them. These scenarios highlight the need for a methodology that can be applied to both the development phase of the intervention and post-hoc analyses, to ensure that adequate data is recorded, and efficiently and systematically scrutinised.

This thesis presents a methodological framework that supports identifying and analysing usage metrics that are meaningful to a specific intervention (Chapter 3). Through analyses of these metrics, the findings provide useful and practical insights for how a specific intervention works, and therefore how other similar interventions may be optimally designed. Together, these types of findings contribute to the digital health field by furthering our understanding about how we can create and meaningfully evaluate interventions that effectively engage target users (see Chapters 4 and 5).

1.3 Thesis structure

The thesis is structured in the three paper format; research underpinning each paper was designed to address the overall aim and specific objectives outlined previously. The three papers contained in this thesis present a novel methodology to structure the process of usage analyses (Chapter 3), and applied examples demonstrating the benefits and insights provided by using this method (Chapters 4 and 5). The research reported in this thesis was carried out between 2016 and 2021. A scoping review of reviews (Chapter 2) was carried out in 2021 to provide an overview of directions in conceptualising and applying usage, and a context for the contribution of this research. An overview of each paper is provided in this chapter, including the specific objectives for each paper and the overall research objectives that the paper meets.

1.4 Individual contributions

Contributions by the supervisory team and me are listed below, with additional author contributions provided in individual paper overviews.

1.4.1 Supervisory team

The supervisory team provided feedback on structure and format for the thesis Introduction, Discussion and Scoping Review. All supervisors are co-authors on the three papers and advised on the development of these (see Chapters 3-5).

Dr Leanne Morrison, primary main supervisor (50%). Secured PhD funding based on the initial research title in collaboration with co-supervisors. Advised on theories and developments in the field of DBCI engagement. Helped shape and format drafts of research papers and thesis.

Prof Lucy Yardley, second/co-supervisor (20%). Developed and trialled PRIMIT and Internet Dr interventions, and advised on their structure, content and theoretical-underpinning. Helped shape the AMUsED framework to reflect existing analysis processes.

Prof Peter Smith, second/co-supervisor (20%). Advised on appropriate statistical analyses and methods for usage analyses.

Dr Mark Weal, second/co-supervisor (10%). Advised on application to different technology and data capture processes. Provided guidance on presenting study processes, intervention structure and results.

1.4.2 My contribution

The thesis Introduction and Discussion chapters, Scoping Review and three papers were conceived and written by me. I created the Scoping Review protocol and carried out the review. The AMUsED framework was designed by me while carrying out preliminary usage analyses. The process evaluation plans and subsequent analyses were designed and undertaken by me. For the Germ Defence study, I was responsible for adapting the intervention, designing the study, seeking ethical approval, and disseminating to the public. Responses to reviews of papers and making required revisions were undertaken by me.

1.5 Overview of scoping review and three papers

1.5.1 Defining and measuring usage of digital behaviour change interventions: a scoping review of reviews

1.5.1.1 Overview of the review

A scoping review of reviews of DBCIs was undertaken, adopting a narrative synthesis approach to explore and understand both developments in the field over time and to evidence the need for systematic usage analyses. Ten reviews were identified that met the inclusion criteria. Review authors agreed that usage analyses are important for providing findings that will inform the development of interventions and better explain positive behaviour change. However, views on the best methods to achieve this varied, with half the authors arguing for a standardised list of summative metrics to enable comparison across interventions (Beintner et al., 2019; Fleming et al., 2018; Nelson, Coston, Cherrington, & Osborn, 2016; Pham, Graham, et al., 2019). Yet the value of summative usage metrics is questioned for their inability to provide precise indicators of intervention engagement (Donkin et al., 2011). Instead, other authors argue for the use of fine-grained usage metrics that represent the specific intervention's structure and features, and the type of interactions with them, as these are more likely to explain how positive behaviour change occurred (Beintner et al., 2019; Koneska, Appelbe, Williamson, & Dodd, 2020; Perski et al., 2017; Smith & Liu, 2020).

1.5.1.2 Paper objectives

- To identify developments in the field of DBCI engagement over the research period.
- To explain the use of terminology within the field and how these are conceptualised in relation to usage.
- To understand how usage has been operationalised and applied in analyses.
- To provide context by which to assess the contribution of the research to the field.

1.5.1.3 Thesis objectives

- To describe concepts currently used to define and analyse usage, and critically discuss how they have been applied to DCBIs, to demonstrate the contribution of the PhD research within a rapidly developing field.

1.5.1.4 Contributions by additional author

Kinga Olexa: Assessed full-text articles for eligibility and final inclusion.

1.5.2 Paper 1 - Supporting Systematic Assessment of Digital Interventions: A Framework for Analysing and Measuring Usage and Engagement Data (AMUsED).

1.5.2.1 Overview of paper

The framework for **A**nalysing and **M**easuring **Us**age and **E**ngagement **D**ata (AMUsED) provides a methodology that helps identify theory-, evidence- and content-based rationales for selecting usage metrics that are meaningful to the intervention (i.e. will help identify how change occurred, and provide insight for improvement and dissemination). The framework can be applied to a range of digital interventions to help structure efficient and systematic usage analyses of large-scale data. The paper includes three checklists to support the following stages: (1) familiarisation with the intervention and its relationship to the captured data, (2) identification of meaningful measures of usage and specifying research questions to guide analyses of usage data, and (3) preparation of datasheets and consideration of available analytical methods with which to examine the data.

1.5.2.2 Paper objectives

- To provide a methodology to structure systematic usage analyses that can be applied both during the development phase to help identify necessary data collection, and after the trial for process evaluations.
- To explain the purpose and benefits of identifying and applying meaningful usage metrics for inferential analyses.
- To support the generation of research questions that will examine usage in relation to underlying theories of the intervention, target outcomes, and by different user groups.
- To prompt consideration of data capture methods and formats during development for compatibility with proposed analytical tools.

1.5.2.3 Thesis objectives

- To explain how large-scale data from digital health interventions can be best utilised to develop our understanding of engagement.
- To develop a method for analysing large-scale data that is systematic and efficient, and produces findings that are meaningful to the intervention and inform our understanding of how theoretically-underpinned interventions are used.
- To enable the consideration of appropriate data analyses and required data collection during the development stage of a digital intervention, and to support post-hoc analyses of pre-collected data.

1.5.2.4 Publication details

Miller, S., Ainsworth, B., Yardley, L., Milton, A., Weal, M., Smith, P., & Morrison, L. (2019). A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. *Journal of Medical Internet Research, 21*(2):e10966. https://doi.org/10.2196/10966

Written between Feb 2016 – March 2018. Submitted to JMIR 4/5/18. Published 15/2/2019.

1.5.2.5 Contributions by additional authors

Ben Ainsworth: Assisted with the concept of effective engagement in application to usage analyses and specifically in formulating appropriate research questions. Supported applying the framework during the development phase of an intervention for Germ Defence.

Alex Milton: Assisted with data extraction and supported early drafts of the framework in relation to the Internet Dr usage analysis.

1.5.3 Paper 2 – How digital interventions support self-care of minor ailments: A process evaluation of the Internet Dr intervention for colds and flu.

1.5.3.1 Overview of paper

The paper presents a process evaluation of data recorded during a randomised controlled trial (RCT) of a digital intervention designed to support users to self-care for colds and flu instead of unnecessarily visiting their GP (Little et al., 2016). The AMUsED framework was applied to structure a post-hoc usage analysis (Miller et al., 2019). By using the framework it was possible to identify and analyse meaningful usage metrics which revealed which parts of the intervention were associated with increased enablement to self-care and which groups of users were more likely to access them.

1.5.3.2 Paper objectives

- To identify measures of usage that are meaningful to the intervention.
- To examine which parts of the intervention worked, for whom and at what time.
- To generate guidance on how to improve the design, implementation and dissemination of the Internet Dr.
- To provide generic guidance on successful design and development for other digital interventions promoting self-care for minor ailments.

1.5.3.3 Thesis objectives

- To enable the consideration of appropriate data analyses and required data collection during the development stage of a digital intervention, and to support post-hoc analyses of pre-collected data.
- To apply and test the methodology to demonstrate how insights may be gained for understanding intervention efficacy, leading to suggestions for improvement, and informing the wider field of engagement with digital interventions.

1.5.3.4 Publication details

Miller, S., Yardley, L., Smith, P., Weal, M., Milton, A., Stuart, B., Little, P., & Morrison, L. (2020). How digital interventions support self-care of minor ailments: A process evaluation of the Internet

Dr intervention for respiratory tract infections. *Journal of Medical Internet Research Preprints,* 24239. https://doi.org/10.2196/preprints.24239

Written between Feb 2016 – Feb 2020. Submitted to JMIR 14/9/2020. Second round of reviews required minor revision, awaiting final decision.

1.5.3.5 Contributions by additional authors

Alex Milton: Assisted with data extraction and first round of Internet Dr usage analysis.

Beth Stuart: Advised on findings from the RCT, understanding and managing RCT data sets, and alternative statistical methods to best examine the data.

Paul Little: Provided feedback on the analysis and paper in relation to the RCT.

1.5.4 Paper 3 – Process evaluations of a web-based intervention to increase handwashing during a pandemic: Moving Germ Defence from a randomised controlled trial to public dissemination

1.5.4.1 Overview of paper

The paper presents two process evaluations of the same intervention used in different contexts: an RCT and open-access to the general public. The intervention supports increased infection prevention and control behaviours in the home to reduce the spread of respiratory tract infections. Having been found to be effective in a large RCT, the intervention was updated as part of this research and disseminated to the general public. The AMUsED framework (Miller et al., 2019) was applied to structure an advance usage analysis plan and to ensure that the appropriate data was recorded during the public dissemination. Comparison of the two contexts provides insights for: the effect of trials on usage; improvements to the intervention; public health handwashing campaigns.

1.5.4.2 Paper objectives

- To compare completion of the online research enrolment procedures across contexts.
- To compare intervention usage by the general public to usage during the RCT.
- To examine whether intentions to wash hands in specific situations were different in the two contexts.

1.5.4.3 Thesis objectives

- To enable the consideration of appropriate data analyses and required data collection during the development stage of a digital intervention, and to support post-hoc analyses of pre-collected data.
- To apply and test the methodology to demonstrate how insights may be gained for understanding intervention efficacy, leading to suggestions for improvement, and informing the wider field of engagement with digital interventions.

1.5.4.4 Publication details

Miller, S., Ainsworth, B., Weal, M., Smith, P., Little, P., Yardley, L., & Morrison, L. (2021). Process evaluations of a web-based intervention to increase handwashing during a pandemic: Moving Germ Defence from a randomised controlled trial to public dissemination. *Journal of Medical Internet Research,* 26104 (forthcoming/in press).

Written between Mar 2019 – Nov 2020. Submitted to JMIR 27/11/2020. Accepted 31/5/21.

1.5.4.5 Contributions by additional authors

Ben Ainsworth: Helped develop the Germ Defence intervention for public dissemination and data capture.

Paul Little: Provided feedback on the analysis and paper in relation to the RCT.

1.6 Impact of COVID-19 on research

Due to the additional demands on researchers, journals have found difficulties in finding reviewers. As a result the Internet Dr process evaluation paper was with a journal for several months until the paper was rejected because they were unable to find reviewers. The paper was subsequently submitted to JMIR. However, due to the relevance of the third paper to the pandemic, the revision and acceptance process was accelerated.

Working on a COVID-19 project and having had long-COVID for five months meant that less time was available to work on the research, and extensions were required to carry out revisions to the last two papers.

1.7 Definitions of terminology used within the thesis

The terminology and approach used in this thesis is in line with the multi-dimensional model of engagement proposed by Perski et al. (2017) as this provides the most all-encompassing definition associated with usage. By including the user’s subjective experience, the concept of overall engagement is placed within the user’s volition, thereby emphasizing this aspect within usage as well (see Scoping Review in Chapter 2). In comparison, the definition for ‘adherence’ covers only usage, and is defined by the research development team as opposed to the user (Beintner et al., 2019; Donkin et al., 2011; Kelders et al., 2012; Sieverink, Kelders, & Gemert-Pijnen, 2017; Smith & Liu, 2020).

The following terms and their definitions are used in this research. They have been selected both for their ability to describe aspects of usage and their underlying theoretical approach.

1.7.1 Analyses

Descriptive analyses: descriptive statistics (e.g. usage metrics for number of users, mean time spent, frequency of logins).

Inferential analyses: bivariate and multivariate analyses (e.g. associations between usage metrics and user characteristics or outcomes).

Usage analysis: analyses of variables representing user interaction with an intervention.

Process evaluation: assesses implementation processes, causal mechanisms in changes in outcomes, and identifies contextual factors associated with variations in outcomes (Craig et al., 2008).

1.7.2 Attrition

Dropout attrition: participants leaving the research process (e.g. not completing measures at follow-up) (Eysenbach, 2006).

Nonusage attrition: no longer using the intervention (Eysenbach, 2006).

1.7.3 Data

Accrued data: collected digitally through research procedures or as part of the intervention. Includes automatically and unobtrusively recorded interactions (e.g. usage data) and user-entered data (e.g. self-report, user characteristics, behavioural outcomes).

Contextual data: data regarding the context around the intervention (e.g. previous findings regarding usage, external factors affecting usage).

Log-data: data collection that records specified interaction with the intervention. This may not include all potential usage metrics as log-data only captures variables that have been selected and programmed into the data capture process.

1.7.4 Engagement

DBCI engagement: based on Perski et al.'s (2017, p.258) definition of comprising of two dimensions "*(1) the extent (e.g. amount, frequency, duration, depth) of usage and (2) a subjective experience characterised by attention, interest and affect*".

Effective engagement: the level of use sufficient to bring about positive behaviour change. This is identified by analysing what type of usage led to positive behaviour change (Yardley et al., 2016).

Effective features: sections of an intervention that are associated with positive outcomes, regardless of the amount or frequency of use (Perski et al., 2017).

1.7.5 Usage

Fine-grained usage metrics: measures of usage that capture interactions that are temporal, contextual, content and user driven (e.g. number of pages viewed on initial login, type of content accessed during period of increased illness, most accessed group/type of content).

Intended usage: usage patterns anticipated by the developer (e.g. accessing components in a specific order or a certain time). The term is used in this context to encourage usage analyses that compare actual and intended usage to uncover unexpected patterns of usage behaviour. This is different from intended usage as defined by adherence whereby users who do not follow the developer's prescribed level of usage are considered to be non-adherent.

Meaningful usage: selecting usage metrics that capture the intervention's architecture and theoretical basis.

Summative usage metrics: measures of usage across the whole sample and intervention e.g. number of logins, number of pages viewed, number of components used, time spent on the intervention.

Usage metrics: variables that measure interaction with a digital intervention (e.g. number of logins, time spent on a page, frequency of data entry, repeat use of a feature)

Chapter 2 Defining and measuring usage of digital behaviour change interventions: a scoping review of reviews

2.1 Introduction

The overall aim of this research was to devise a methodology to guide and help structure usage analyses, and to apply this to large, secondary datasets collected during RCTs to better understand how the interventions were effective. Typically, a review would be carried out at the start of the research period to identify, describe and compare existing frameworks or protocols for analysing DBCI usage so that the findings may inform the research. However, at the start of the research period, to the best of my knowledge, no other methodologies were available that addressed usage analyses. On that basis it was considered that a review would not be able to address the research question at that time.

The lack of usage analysis methodologies available in 2016 highlights the need for one in this area. However, DBCI evaluation is an emerging field, and other research teams have also recognised the importance of conceptualising and measuring usage for understanding intervention efficacy (e.g. Department of Behavioural Science & Health, University College London; Department of Psychology, Health and Technology, University of Twente). As a result there has been an increase in research on this over the past five years. Yet agreement has not been reached, and usage is conceptualised, measured and reported using divergent methods (Perski et al., 2017). In addition, terminology is often applied interchangeably when referring to different concepts, such as ‘engagement’ and ‘usage’. Therefore, it was considered that undertaking a scoping review at the end of the research period would be beneficial for providing an overview of the diverse approaches, terminology and definitions applied across the field, how they have developed and what still remains unresolved. This would also provide a context against which the contribution of the current research can be considered and assessed.

The aim of the review was to describe how usage has been conceptualised and applied over the last two decades. As the evaluation of DBCI usage is still an emerging area and the aim is to clarify concepts within that, a scoping review was considered appropriate (Munn et al, 2018). The majority of published usage analyses focus on reporting usage for one intervention. Therefore, a review of reviews was considered more appropriate to capture changes in the field over time and the emergence of overarching concepts and methods for usage analysis.

The research questions were:

- What concepts are used to define usage of DBCIs?
- How are these concepts measured?

2.2 Methods

The scoping review was structured by Arkey and O'Malleys (2005) methodological framework, with definitions and methodological stages also informed by additional guidance (Colquhoun et al., 2014; Levac, Colquhoun, & O'Brien, 2010; Peters et al., 2015). The results are reported using a narrative synthesis approach to bring together the concepts and how they have been defined, and explain how these have changed and evolved over time (Popay et al., 2006).

2.2.1 Data sources and search strategy

Searches of PsychINFO, Web of Science and MEDLINE electronic databases were carried out. Five categories of search terms were devised so that all searches included one term from every category as follows:

1) A term related to usage included in the title (i.e. usage, engagement, data, attrition, or adherence) in the title.
2) The term "review" in the title.
3) The term "health" in topic (i.e. title, abstract, keywords).
4) The term "intervention" in topic.
5) A term related to digital in the topic (e.g. internet, online, mhealth, ehealth).

Full details are available in the review protocol, Appendix A.

2.2.2 Eligibility criteria

Articles were screened for inclusion based on the following criteria. 1) Reviews had been published in a journal between January 2000 and December 2020, and had been peer-reviewed. January 2000 was chosen as few digital interventions existed before then, making reviews of findings impracticable. 2) The studies selected and reported in the reviews were digital health behaviour change interventions. 3) Reviews reported, defined or categorised objective, quantifiable usage metrics representing usage of a DBCI. 4) The findings reported were applicable to web- and app-based interventions as these platforms are able to produce a broad range of variables that provide in-depth and fine-grained detail of intervention usage. 5) Reviews were published in English.

2.2.3 Review selection

The review selection was carried out by me (SM), with a voluntary research assistant, Kinga Olexa (KO) who also decided independently on the eligibility of full-text articles for inclusion. First screening was based on publication titles; reviews were excluded if they clearly did not meet the criteria (i.e. were not reviews, were in areas other than health, interventions were not digital, engagement was with services, treatments or community, see Figure 1). The reviews were downloaded, merged and duplicates removed using Rayyan QCRI (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016). These were further screened based on abstracts. In a blind assessment of full-texts, SM and KO agreed on all reviews for final inclusion.

2.2.4 Data extraction and analysis

A data extraction table was developed and applied by the first author to extract the following items:

- General information: review title, authors, journal, year and first author location
- Type of review, target condition and number of studies reported
- Review objective
- Definitions applied or developed
- Usage metrics collected/reported
- Recommendations

A narrative synthesis approach was used to analyse the data extracted from the reviews (Popay et al., 2006). The reviews were initially organised chronologically to assess changes over time. Textual descriptions of each of the reviews were produced to briefly describe an overview of the contents and aide familiarity with them. Reviews were identified as being more ‘weighty’ if their content was more substantial leading to greater data extraction (Popay et al., 2006). Relationships between aims and recommendations were explored and reviews grouped accordingly. The definitions provided in reviews were analysed and further groupings made based on the main concept they addressed. The usage metrics identified in the reviews were tabulated to enable comparison of the findings.

2.3 Results

2.3.1 Review selection

Ten reviews were selected for final inclusion for the scoping review (Beintner et al., 2019; Donkin et al., 2011; Fleming et al., 2018; Kelders et al., 2012; Koneska et al., 2020; Nelson et al., 2016; Perski et al., 2017; Pham, Graham, et al., 2019; Sieverink, Kelders, & Gemert-Pijnen, 2017; Smith & Liu, 2020) (see Figure 1).

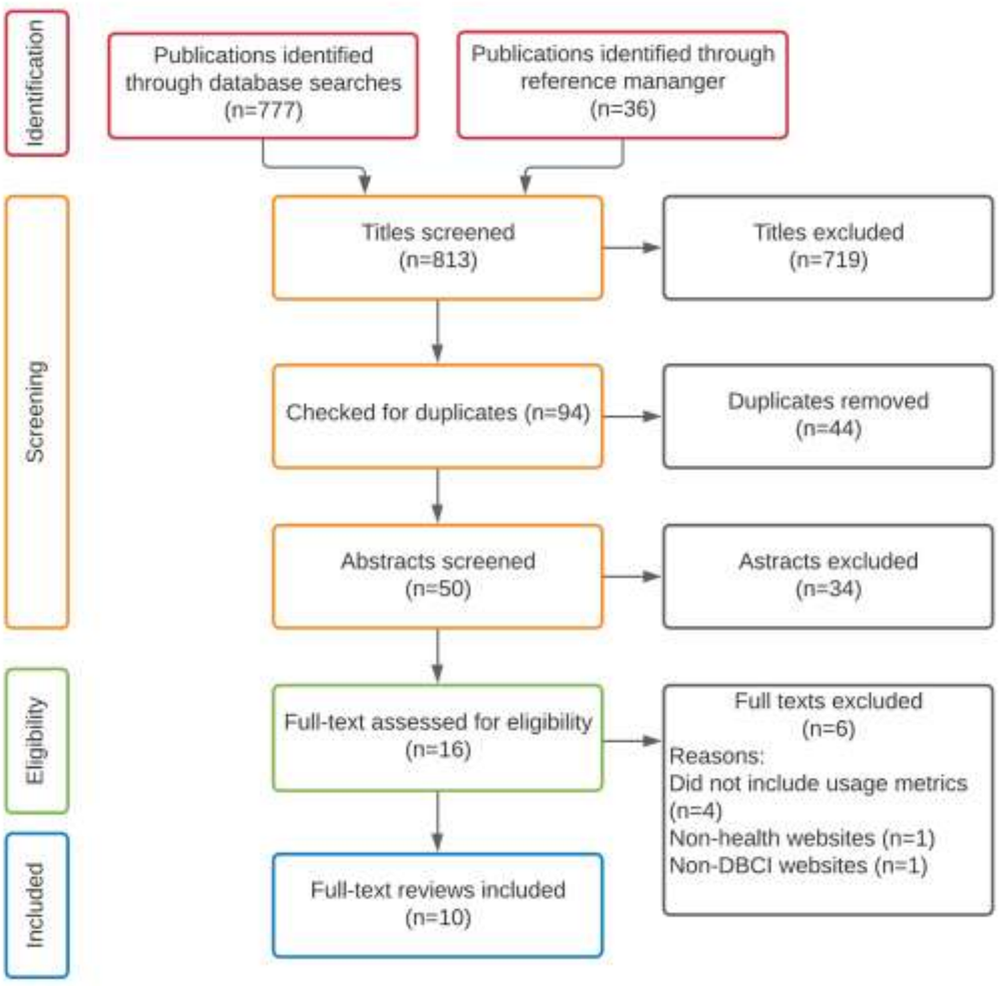


Figure 1. Scoping review selection process

2.3.2 Review characteristics

By charting the reviews chronologically it is apparent that eight out of the ten were published in the last five years (see Figure 2). First authors were located in North America, Northern Europe or Australasia. Analysis of the relationships between the reviews' aims and recommendations revealed two groups. The first group consists of six reviews that had been carried out to identify usage metrics previously applied in studies (Beintner et al., 2019; Fleming et al., 2018; Koneska et al., 2020; Nelson et al., 2016; Pham, Graham, et al., 2019; Smith & Liu, 2020). Three of these reviews focused on interventions for specific conditions, with the aim of understanding what type of usage was effective in supporting positive outcomes for that condition (Fleming et al., 2018; Nelson et al., 2016; Smith & Liu, 2020). Within this first group, the authors found that reported usage metrics varied widely and were inconsistently applied. As this made their objectives harder to achieve, the reviewers called for more systematic or uniform methods of measuring and reporting DBCIs (Beintner et al., 2019; Fleming et al., 2018; Koneska et al., 2020; Nelson et al., 2016; Pham, Graham, et al., 2019; Smith & Liu, 2020). The four remaining reviews fell in to a second group with different objectives: the researchers wanted to understand and explain how usage metrics inform the concepts of adherence and engagement (Donkin et al., 2011; Kelders et al., 2012; Perski et al., 2017; Sieverink, Kelders, & Gemert-Pijnen, 2017). When extracting data, these reviews produced longer textual descriptions and were identified as having more weight.

Analysis of the definitions provided in the reviews revealed three main concepts: adherence, engagement and usage. Of the ten reviews included, five focus on the concept of adherence to DBCIs (Beintner et al., 2019; Donkin et al., 2011; Kelders et al., 2012; Sieverink, Kelders, & Gemert-Pijnen, 2017; Smith & Liu, 2020). Four reviews claim to examine the concept of ‘engagement’, but for two of these engagement is synonymous with usage (Fleming et al., 2018; Nelson et al, 2016), leaving two reviews that focus on engagement (Perski et al., 2017; Pham, Graham et al., 2019) and three on usage metrics (Fleming et al., 2018; Koneska et al., 2020; Nelson et al., 2016) (see Figure 2).

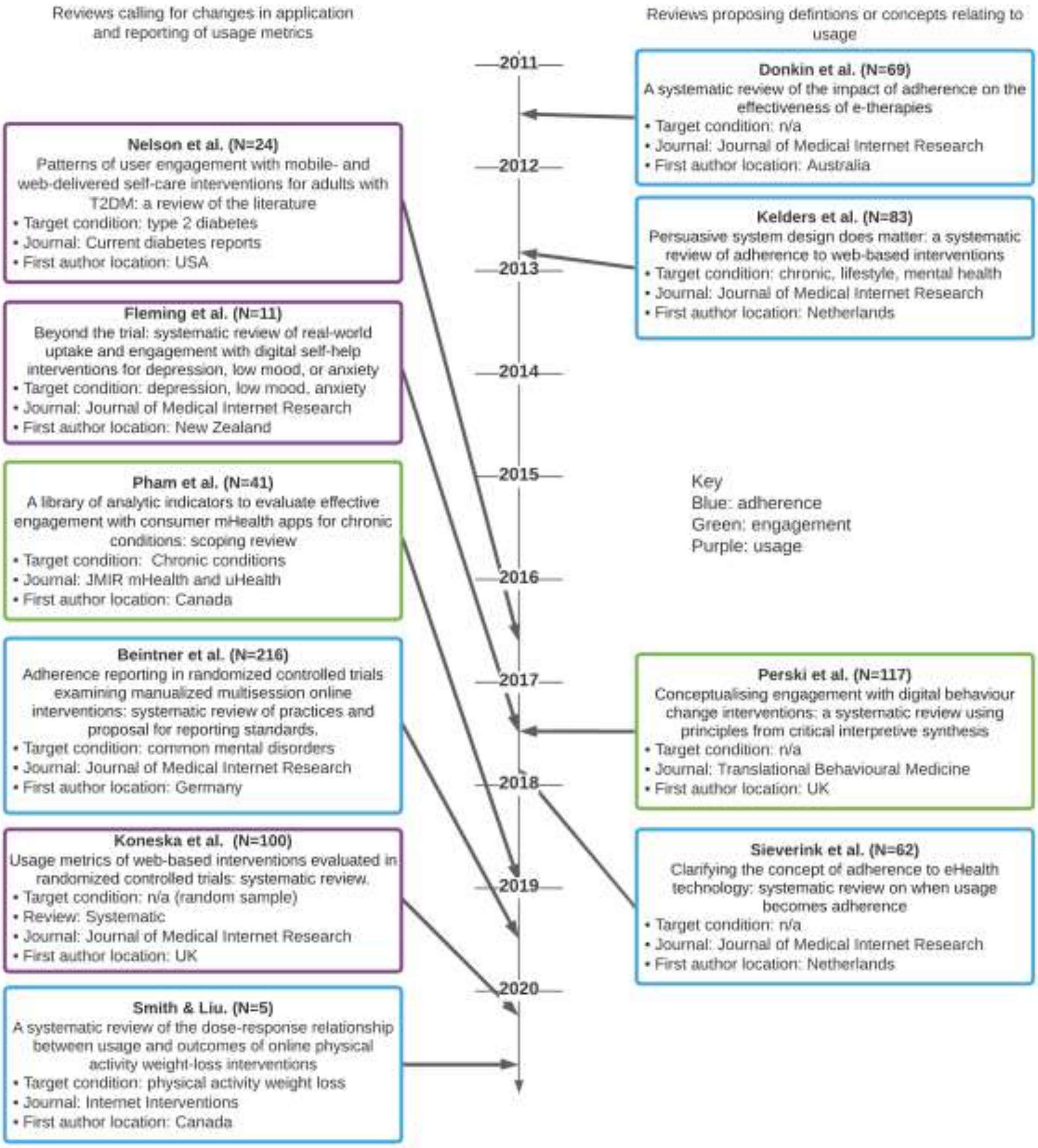


Figure 2. Reviews arranged chronologically and grouped by aims/recommendations.

2.3.3 Definitions

As identified in study characteristics, reviews concerned with providing and advancing current definitions provided new conceptualisations. The remaining reviews focused on identifying common usage metrics and therefore provided fewer novel insights for definitions.

2.3.3.1 Adherence

Half of the reviews are based around the concept of adherence (Beintner et al., 2019; Donkin et al., 2011; Kelders et al., 2012; Sieverink, Kelders, & Gemert-Pijnen, 2017; Smith & Liu, 2020).

Applications of adherence in these reviews may be categorised as: the development of the term from the medical application; adherence in relation to DBCI usage; and ways in which adherence may be captured and measured.

2.3.3.1.1 Medical adherence

Medical adherence is defined as the extent that users follow treatment (e.g. medication, advice) as advised by a health care professional (Donkin et al., 2011; Kelders et al., 2012). This is conceptualised through a dose-response relationship whereby the more that a patient follows their recommended treatment (i.e. dose) the better the associated outcomes will be (i.e. response; Donkin et al., 2011). This means that a certain level of medication is necessary to reach the required outcome (Donkin et al., 2011).

2.3.3.1.2 Defining usage adherence

Developed from medical adherence, adherence to DBCIs is based around the amount an intervention is used, with usage as the ‘dose’ in the dose-response relationship (Sieverink, Kelders, & Gemert-Pijnen, 2017). Usage adherence is described as the extent that users interact with a DBCI as deemed necessary by the intervention developers to achieve positive outcomes (Beintner et al., 2019; Donkin et al., 2011; Kelders et al., 2012; Sieverink, Kelders, & Gemert-Pijnen, 2017; Smith & Liu, 2020). Beintner et al. (2019) propose that when applied to DBCIs, ‘adherence’ is an umbrella term that captures non-usage attrition, engagement, user retention, persistence, exposure, and intervention usage.

2.3.3.1.3 Operationalising adherence

Within the reviews focusing on adherence, three different types of metrics are used to measure adherence. Donkin et al. (2011), Beintner et al. (2019) and Smith and Liu (2020) use measures of user interaction with the intervention as adherence metrics. Kelders et al. (2012) propose that by comparing actual usage to intended usage, a percentage of users that interacted with the intervention as intended may be constructed to capture adherence. Intended usage is defined as the type of interaction that the intervention development team believe is necessary in order to achieve positive outcomes (also called optimal dose, intervention dose; Kelders et al., 2012). For their analyses, Kelders et al. (2012) use increased adherence as the positive outcome. Sieverink, Kelders, and Gemert-Pijnen (2017) categorised studies in to three groups depending on how adherence is conceptualised: 1) the more usage the better; 2) levels of intended usage are provided without a rationale; 3) intended usage is justified using theory or evidence. Over half the studies in their review define adherence as ‘the more usage the better’ or using the whole intervention (Sieverink, Kelders, & Gemert-Pijnen, 2017).

2.3.3.2 Engagement

Although the term ‘engagement’ is the focus of four of the reviews, authors for two of these define and apply ‘engagement’ as being the same as ‘usage’ (Fleming et al., 2018; Nelson et al., 2016). Of the two remaining, Pham, Graham et al. base (2019) their review around the concept of effective engagement as previously defined by Yardley et al. (2016), but they do not expand on the original definition. Perski et al. (2017) provide a new definition for engagement based on their review which draws from studies within computer, behavioural and human computer interaction sciences. They define engagement with DBCIs as comprising of two dimensions: “*(1) the extent (e.g. amount, frequency, duration, depth) of usage and (2) a subjective experience characterised by attention, interest and affect*” (Perski et al., 2017, p.258).

2.3.3.3 Usage

All authors agree that usage metrics record the actual, objective interaction between the user and the intervention recorded as ‘log-data’. Every review collected the usage metrics identified in their selected studies with the aim of understanding how that data may provide insight in to achieving better outcomes.

2.3.3.3.1 Operationalising usage

Usage metrics reported in the reviews are summarised in Table 1. Pham, Graham et al. (2019) and Perski et al. (2017) use previous categorisations of usage metrics to identify amount, breadth, duration, frequency, and depth of interaction (Couper et al., 2010; Danaher et al., 2006; O’Brien & Toms, 2008). However, the authors apply the definitions to different usage metrics. The two usage metrics most commonly reported in the reviews were summative measures of total number of logins and the number of modules/components completed.

Table 1. Usage metrics identified through the scoping reviews of DBCIs.

Usage metrics identified	Beintner et al	Donkin et al	Fleming et al*	Kelders et al*	Koneska et al	Nelson et al	Perski et al	Pham et al	Sieverink et al	Smith et al	Total
Number of logins	✓	✓	✓	✓	✓	✓	✓	✓ amount	✓	✓	10
Number of modules completed	✓	✓	✓	✓	✓	✓	✓	✓ depth	✓	✓	10

Time spent	✓	✓			✓	✓	✓ amount/ breadth	✓ duration	✓	✓	8
Number of features used		✓			✓	✓	✓ depth	✓ breadth	✓	✓	7
Data input	✓	✓	✓		✓	✓		✓ depth			6
Number of pages viewed		✓			✓		✓	✓ breadth	✓		5
Login frequency			✓	✓		✓	✓ frequency	✓ amount		✓	5
Intervention completed	✓		✓		✓	✓					4
Number of exercises completed		✓							✓		2
Time period of use			✓				✓ duration		✓		2
Responding to messages	✓										1
Login duration										✓	1
Repeat content use										✓	1
Point of attrition	✓										1
Percentage of content accessed										✓	1

* Authors created their own measures of usage/adherence by grouping several variables together or re-classifying them.

2.4 Discussion

2.4.1 Review characteristics

By charting the reviews chronologically it is evident that interest in measuring and defining usage has grown over the last five years, with eight of the ten reviews occurring in that period. Although this is indicative of the increase in studies now reporting usage analyses for digital behaviour change interventions, the number of studies identified as having reported sufficient usage metrics (both in the amount available and with the necessary detail) to meet the inclusion criteria is still low in some reviews (Fleming et al., 2018; Nelson et al., 2016; Smith & Liu, 2020). This suggests that whilst this area is expanding, the number and type of usage metrics reported are still limited.

2.4.2 Definitions

Four reviews were identified as proposing definitions or concepts relating to engagement. Of these, three looked specifically at adherence (Donkin et al., 2011; Kelders et al., 2012; Sieverink, Kelders, & Gemert-Pijnen, 2017) and one at engagement (Perski et al., 2017). Through the definitions these reviews provide, it is possible to compare and contrast these two concepts and better understand their influence on usage analyses.

2.4.2.1 Comparing and contrasting adherence and engagement

The concept of adherence to a digital intervention refers solely to usage. Grounded in medical adherence, usage adherence is based around the idea that the intervention development team decide the amount of usage necessary to bring about positive behaviour change (Donkin et al., 2011; Kelders et al., 2012). This decision may be based on theory, previous evidence from other interventions or the behaviour itself. Once decided upon, the intervention is designed with the intention that it should be used in full, and that less usage will reduce positive outcomes. On this basis, Kelders et al. (2012) treated the amount an intervention had been used as if it were the outcome measure, removing the focus from behaviour change. This led to the conclusion that intervention content should be designed with the goal of encouraging increased usage (Kelders et al., 2012; Nelson et al., 2016). However, constructing measures of adherence and analysing them in place of behavioural outcomes is problematic. Adherence to an intervention is not necessarily the same as adherence to the target behaviour (Sieverink, Kelders, & Gemert-Pijnen, 2017). Users may choose to disengage from an intervention for a variety of reasons, including having reached their personal goals (Sieverink, Kelders, & Gemert-Pijnen, 2017). In addition, Sieverink, Kelders, and Gemert-Pijnen (2017) found that the majority of studies lacked clear rationales for their levels of intended usage, and instead relied upon ‘the more usage the better’. Without an explanation for why additional usage is beneficial, replacing outcomes with measures of adherence may aid comparison between interventions, but will not provide a better understanding of how an intervention was used and when it was effective.

In comparison, the concept of engagement is characterised as both the user's subjective experience and the extent to which an intervention is used, with these two aspects interacting (Perski et al., 2017). This places the focus on the user and their decision to interact with the intervention, as opposed to expected or intended usage levels as defined by the intervention developers. Perski et al. (2017) propose that while intervention content does influence engagement, other factors such as personal characteristics and environmental contexts are also important. Mixed methods are advised to capture and explore these various aspects of engagement: usage analyses to examine users' interaction with an intervention, and qualitative or self-report methods to understand their experience of that interaction (Perski et al., 2017; Pham, Graham et al., 2019). This process informs the usability and acceptability of the intervention, whilst ensuring that the content is also motivating, convincing and informative, and supports the user in positive behaviour change (Yardley et al., 2015).

In contrast to the idea of 'the more usage the better', Pham, Graham et al. (2019) and Fleming et al. (2018) propose that brief usage may still bring about positive changes. Pham, Graham et al. (2019) focus on the concept of effective engagement: that it is possible to identify the minimum type or level of intervention usage that achieves intended outcomes (Yardley et al., 2016). For their review they compile a list of previously reported usage metrics, with the recommendation that application of these in the future will aid identification of effective engagement (Pham, Graham, et al., 2019). Pham, Graham et al. (2019) concluded that applying usage metrics that represent the content and context of the intervention, and analysing these with the specific intended outcomes is most likely to capture effective engagement. Supplementing this definition, Perski et al. (2017) introduce the concept of 'effective features' which suggests that certain intervention features may be associated with increases in positive outcomes. They also emphasise that this may not necessarily be the most used parts of the intervention, for example, people may interact with data entry features on a daily basis, but components providing support or problem solving that are accessed less frequently may provide greater support for positive outcomes (Perski et al., 2017).

The most recent review addressing definitions by Sieverink, Kelders, and Gemert-Pijnen (2017) also introduces a new concept within adherence of a 'threshold for intended usage'. The concept was developed to describe studies where users did not adhere to the intervention, but still experienced positive changes in the target behaviour. Sieverink, Kelders, and Gemert-Pijnen

(2017) define the threshold for intended usage as the minimum use necessary to experience changes to the target behaviour. Although this might appear indistinguishable from the concept of effective engagement, they are underpinned by different theoretical approaches. Effective engagement is defined by the users, and captured using qualitative and quantitative data to understand their interaction and experience with an intervention. The aim is then to shape the intervention using that information to make it more accessible and easy to use where possible. Although the definition for a threshold for intended usage may be shaped by feedback from users, the intervention research team hold responsibility for deciding the minimum level or type of usage expected, as opposed to the user. Sieverink, Kelders, and Gemert-Pijnen (2017) suggest that the practical application for thresholds for usage is that they allow the researchers to measure adherence taking this in to consideration.

2.4.3 Reporting of usage metrics

Tabulation of the usage metrics from all reviews over the decade revealed that the number of times users logged in to an intervention and the number of modules completed were operationalised as usage metrics in all ten reviews. As highlighted in several reviews, these metrics represent summative usage measures, but do not provide much insight for the reader as to how the intervention was used in a way that may change behaviour (Beintner et al., 2019; Koneska et al., 2020; Smith & Liu, 2020). The lack of clarity in what the metrics are capturing is highlighted unintentionally by Perski et al. (2017) and Pham, Graham et al. (2019). These authors have used the categories of amount, breadth, duration, frequency, and depth to define some of the occurring usage metrics. Both authors cite Couper et al. (2010) as having developed these categories, yet it is apparent that the descriptions provided are open to different interpretations, making them less useful in application. Calls are made instead for usage metrics to be selected that capture interaction with specific components, the quality of the interaction, and are supplied with a clear rationale for why the metrics represent intervention interaction that supports behaviour change (Beintner et al., 2019; Koneska et al., 2020; Smith & Liu, 2020).

Further calls for caution in the use of summative measures are also made by the authors focusing on definitions associated with intervention usage. In their review in 2011, Donkin et al. (2011) argue that summative usage metrics may actually be capturing user characteristics that lead to increased engagement, such as motivation or self-efficacy. By utilising fine-grained metrics that capture detailed interaction with an intervention it is possible to explain how change occurred (Donkin et al., 2011). Donkin et al. (2011), Perski et al. (2017) and Sieverink, Kelders, and Gemert-Pijnen (2017) propose that the choice of usage metrics should take in to consideration several factors: variations in users (within and across individuals, over time, user needs and intended outcomes); the intervention (multiple features, different patterns of usage); the target behaviour/condition. These authors emphasise the importance of considering a spread of usage metrics by which to analyse intervention outcomes. The findings in the review by Pham, Graham et al. published in 2019 suggest that these concerns have been influential, as the studies included in their review had moved away from summative measures and utilised more content relevant usage metrics instead.

Reviews that were conducted with the aim of identifying appropriate usage metrics and content for interventions for specific health conditions/behaviours found that the usage metrics reported were inconsistent (Beintner et al., 2019; Fleming et al., 2018; Nelson et al., 2016). This was problematic as they were trying to make comparisons across intervention, so the authors argue for more systematic reporting that includes a set list of both summative and fine-grained metrics (Beintner et al., 2019; Fleming et al., 2018; Nelson et al., 2016). Half of the reviews also highlight the importance of inferential analyses to establish associations between types of usage and behaviour in order to better understand how the intervention worked (Donkin et al., 2011; Nelson et al., 2016; Perski et al., 2017; Pham, Graham et al., 2019; Sieverink, Kelders, & Gemert-Pijnen, 2017). However, with multiple usage metrics available, carrying out analyses on all of the measures will become onerous, and potentially find associations that do not provide insight or meaning (Danaher et al., 2006). In addition, this would not overcome the problem of variability in usage metrics when comparing across interventions, as Perski et al. (2017) found that they were unable to draw conclusions on the relationship between engagement and outcomes due to the multiple ways in which usage had been defined. Analyses driven by research questions are more likely to provide information about the underlying mechanisms that lead to behaviour change, compared to analyses driven by the availability of the metrics (Sieverink, Kelders, Poel, et al., 2017; Yardley et al., 2016). Sieverink, Kelders, and Gemert-Pijnen (2017) found in their review that studies where researchers had clearly defined intended usage, employed less usage metrics as they were able to adopt a more focused approach. Although this approach would not facilitate easier comparison across interventions, for researchers looking to develop interventions for specific conditions, a literature review would highlight the types of content and interactions that are effective, providing a much better idea of how to construct an intervention.

2.4.4 Conclusion

This scoping review describes how definitions and measurement of DBCI usage has developed over the past decade. Only two reviews were identified that had been carried out prior to the start of this research, demonstrating that this was, and still is, an emerging field of research. The lack of agreement on terminology, definitions and measurement is apparent from their varied application within the reviews. However, through the scoping review it was possible to determine two conceptualisations of usage and their underpinning approaches: adherence and engagement. Although the concept of engagement was addressed by fewer reviews, the definition proposed by Perski et al. (2017) has been widely cited. By including and distinguishing behaviour and experience, the role of usage is clearly explained within their definition of engagement. The measurement of usage across the reviews showed that the majority of studies have been reporting summative measures of usage. Whilst some authors felt that this should be a standardised process, others made strong arguments for reporting more fine-grained metrics that examine an intervention’s structure, different user groups and target behaviour.

The scoping review was also intended to provide context for this research. Although the research field has expanded rapidly, the findings in the review continue to support the need for a method to help structure usage analyses that includes a rationale for selecting meaningful measures of usage. By defining and comparing the concepts of adherence and engagement, the research can be positioned conceptually. In Perski et al.’s (2017) definition, usage is considered to be a dimension of engagement alongside the user’s subjective experience, placing usage as being motivated by the user. The process evaluations within this research are based upon this perspective (Chapters 4 and 5). In line with engagement, suggested intervention improvements aim to meet users’ expectations where possible, or encourage and motivate the user to continue when necessary. This is in contrast to adherence, where interventions focus on changing users preferred level or type of interaction to meet researchers’ expectations.

Chapter 3 Paper 1: A Framework for Analysing and Measuring Usage and Engagement Data (AMUSED) in Digital Interventions: Viewpoint

Miller, S., Ainsworth, B., Yardley, L., Milton, A., Weal, M., Smith, P., & Morrison, L. (2019). A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. *Journal of Medical Internet Research, 21*(2):e10966

3.1 Abstract

Introduction: Trials of digital behaviour change interventions (DBCIs) can yield extensive, in-depth usage data, yet usage analyses tend to focus on broad descriptive summaries of how an intervention has been used by the whole sample. This paper proposes a novel framework to guide systematic, fine-grained usage analyses that better enables understanding of how an intervention works, when and for whom.

Framework for Analysing and Measuring Usage and Engagement Data description: The framework comprises three stages to assist: 1) familiarisation with the intervention and its relationship to the captured data; 2) identification of meaningful measures of usage and specifying research questions to guide systematic analyses of usage data; 3) preparation of datasheets, and consideration of available analytical methods with which to examine the data.

Framework application: The framework can be applied to inform data capture during the development of a DBCI and/or in the analysis of data after the completion of an evaluation trial. We will demonstrate how the framework shaped preparation and aided efficient data capture for a DBCI to lower transmission of cold and flu viruses in the home, and informed a systematic in-depth analysis of usage data collected from a separate DBCI designed to promote self-management of colds and flu.

Conclusions: The AMUsED framework guides systematic and efficient in-depth usage analyses that will support standardised reporting with transparent and replicable findings. These detailed findings may also enable examination of what constitutes effective engagement with particular interventions.

3.2 Introduction

Digital behaviour change interventions (DBCIs) are intended to support positive change in a range of health-related outcomes, including psychological, behavioural, educational, social and environmental (Hamine, Gerth-Guyette, Faulx, Green, & Ginsburg, 2015; Newman, Szkodny, Llera, & Przeworski, 2011; Webb, Joseph, Yardley, & Michie, 2010). They may be delivered using any digital device (e.g. phone, computer) making them cost effective for providers (Elbert et al., 2014; Solomon, Proudfoot, Clarke, & Christensen, 2015). Trialling a DBCI can yield complex, large-scale datasets containing detailed usage data. If analysed appropriately, this data is able to provide invaluable detail on how users interact with the intervention, and inform our understanding of engagement. Measuring DBCI engagement has been described as a multidimensional concept, including the extent to which an intervention is used (e.g. amount, frequency, duration), and the subjective experience of the user as characterised by attention, affect and interest (Perski et al., 2017). As a key element of engagement, in-depth and consistently applied usage analyses are capable of providing invaluable insight in to the field of engagement with DBCIs.

Usage analyses frequently examine the extent to which an intervention is used by the whole sample, utilising variables such as the number of times users logged in, total time spent on an intervention, or number of pages viewed (Nelson et al., 2016). These broad level analyses do not always take advantage of the detailed and comprehensive data available, and they frequently assume that greater amounts of usage are indicative of higher levels of interaction that lead to increased changes in target behaviour (Yardley et al., 2016). Harnessing the full range of data can instead enable more informative usage variables to be computed or combined which may answer specific research questions about patterns of usage, i.e. who the intervention was used by and how it was used (Sheeran, Klein, & Rothman, 2017; Yardley et al., 2016). Recent interest in ‘effective engagement’ considers these individual patterns of usage and the minimum level of engagement necessary for changes in target behaviour to occur, including variation across individuals (Ainsworth et al., 2016; Michie et al., 2017; Morrison et al., 2018; Yardley et al., 2016). Effective engagement is defined as sufficient engagement with the intervention to achieve intended outcomes (Yardley et al., 2016). For example, a DBCI designed to lower the transmission of cold and flu provided 4 sessions of content, requiring the intervention to be accessed on 4 separate occasions. However, analysis of usage data, reported behaviour and incidences of illnesses revealed that using the first session alone facilitated the required change in behaviour to increase positive outcomes (Ainsworth et al., 2016). Alternatively, effective engagement may be context-dependent whereby viewing a specific content component, in a certain order, or at an appropriate time, is the minimum threshold necessary for change (Ainsworth et al., 2016; Morrison et al., 2018; Yardley et al., 2016).

During the planning of a DBCI, processes such as logic models and guiding principles may be used to structure the theoretical underpinning and associated content for the intervention (Band et al., 2017; Mohr, Schueller, Montague, Burns, & Rashidi, 2014; Yardley et al., 2015). These techniques help identify behavioural determinants (e.g. beliefs associated with the target behaviour) which may be important in influencing the target behaviour, for example low confidence to manage symptoms, or perceived barriers to performing a specific behaviour. In order to influence favourably these behavioural determinants, content containing behaviour change techniques (BCTs) such as goal-setting, feedback on behaviour, or habit formation are incorporated into the intervention (Michie et al., 2013). However, by performing analyses which focus solely on broad usage patterns across the whole sample, the opportunity is missed to understand how specific intervention content (e.g. BCTs) are used by sub-groups with particular characteristics (i.e. behavioural determinants), and the extent to which there is a relationship to the target behaviour (Ainsworth et al., 2016; Perski et al., 2017). Devising a plan of analysis to answer these questions using the fine-grained data often available from DBCIs enables us to examine the constructs of the logic model and further our understanding of the mechanisms of action underlying successful behaviour change (Michie et al., 2016; Michie, Johnston, Francis, Hardeman, & Eccles, 2008)..

Analysing usage metrics to better understand engagement has been proposed for some time, with the 'Law of Attrition' being one of the first theories to draw attention to the benefits of examining usage data in this way (Eysenbach, 2005). The importance of the type of content viewed as well as the amount has also been acknowledged (Danaher et al., 2006). More recently researchers have advocated using complex log-data from DBCIs to further our understanding of engagement (Ainsworth et al., 2016; Morrison et al., 2018), and examine relationships between usage, participant characteristics and health outcomes (Paz Castro, Haug, Filler, Kowatsch, & Schaub, 2017). However, the importance of providing consistently reported findings that will enable comparison of usage across different DBCIs has also been highlighted (Perski et al., 2017; Yardley et al., 2016). Existing guidelines encourage precise and standardised reporting for general analyses of digital interventions (Craig et al., 2008; Eysenbach & CONSORT-EHEALTH Group, 2011). The challenge of undertaking efficient and systematic analysis of large data sets without the guidance of a framework is already acknowledged: Sieverink, Kelders, Poel et al. (2017) detailed the importance of using research questions to guide analysis of log-data; and Taki et al. (2017) demonstrated how different usage metrics can inform our understanding of engagement. However, systematic reviews suggest that these types of analysis of usage data are not yet routinely undertaken (Nelson et al., 2016; Sieverink, Kelders, & Gemert-Pijnen, 2017). This may be due to the absence of a framework that contains comprehensive checklists combining both the systematic breakdown of usage data and the formulation of research questions to structure usage analyses of DBCIs. In addition, without prior identification of necessary data capture processes, the final usage data collected may be unable to answer the research questions posed.

This paper proposes a novel framework to structure the process of analysing usage associated with a DBCI by 1) drawing together potential measures of usage and identifying which are meaningful to the intervention, 2) generating specific research questions to act as testable hypotheses, and 3) supporting data preparation and selection of methods for analysis. Specifically, the framework for Analysing and Measuring Usage and Engagement Data (AMUsED) can encourage the collection and/or extraction of data that will explain who used which parts of the intervention at what time, and whether that was associated with positive outcomes. The framework focuses on usage as a key component of engagement, but does not aim to encompass all aspects of engagement. Nonetheless, the examination and analyses of usage data, using the framework, can move toward the identification of what constitutes effective engagement. In addition, the framework offers an approach to DBCI data analysis which can be applied both before and after data collection.

When used during intervention development, the AMUsED framework aids development teams to compile an ‘a priori’ analysis plan for use after data collection. This allows the opportunity to evaluate whether all necessary data will be collected, and whether this is in a suitable format for analysis at a later date. This is particularly pertinent for interventions that are developed with external partners who may be unaware of the theoretically-based elements of the intervention and their implications for analyses. When applied after data collection, the framework is especially useful for general orientation when a researcher is unfamiliar with the intervention, or no advance plan of analysis is available. Using the framework helps focus exploratory usage analyses on addressing the theory underpinning an intervention and the plausible mechanisms of action on target outcomes, aiding more scientifically rigorous analyses. Should an analysis plan be available, the framework facilitates a review to ensure that the plan is still appropriate, and aids revision where necessary.

3.3 Development of the framework

The AMUsED framework was initially developed as a means to systematically and rigorously analyse post-hoc usage data collected during DBCI trials. The first author was tasked with analyzing usage data from Internet Dr, a successfully trialled DBCI (see case study 2 below). This task was challenged by the author’s unfamiliarity with the intervention, depth and complexity of data collected, and the absence of an existing framework to provide step-by-step guidance on approaching a usage analysis. Stages 1 and 2 of the framework were developed alongside the process of understanding and beginning analyses of usage data collected from Internet Dr. An early version of the framework was presented to a multi-disciplinary DBCI development team with experience across health psychology, primary care and statistics. The framework was then refined based on the team's input and experiences of applying the framework to their own usage analyses. The value of having a systematic process through which to consider data collection during the development phase of a DBCI was subsequently noted. The framework was then expanded and applied to the amendment of a second intervention, Germ Defence (see case study 1 below). Following this, the framework was presented to the wider scientific community at a national conference in the UK. Here, the value of using the structure provided by the framework to support collaboration between social scientists and software development companies and identification of necessary data collection processes was recognized. The framework was then shaped further to provide equal weight to both a priori and post-hoc analysis needs.

3.4 AMUsED description

3.4.1 Overview

The framework is presented in three stages: 1) familiarisation with available data sets, 2) selecting meaningful measures of usage and generating research questions, 3) preparation for analysis. Each stage is available in checklist format, with generic questions acting as prompts for the researcher to consider in the context of their own specific intervention (see Section 3.7 at end of paper). It is anticipated that use of the three stages will be iterative depending upon whether the framework is being applied in advance of or after data collection (see Figure 3). For example, when considering appropriate analytical software (stage 3) during the development phase of an intervention, it may be necessary to reformat how data is recorded to ensure compatibility. Alternatively, analyses of collected data may reveal unexpected patterns of usage, such as repeated visits to a component of content, from which new exploratory research questions can then be generated (stage 2).

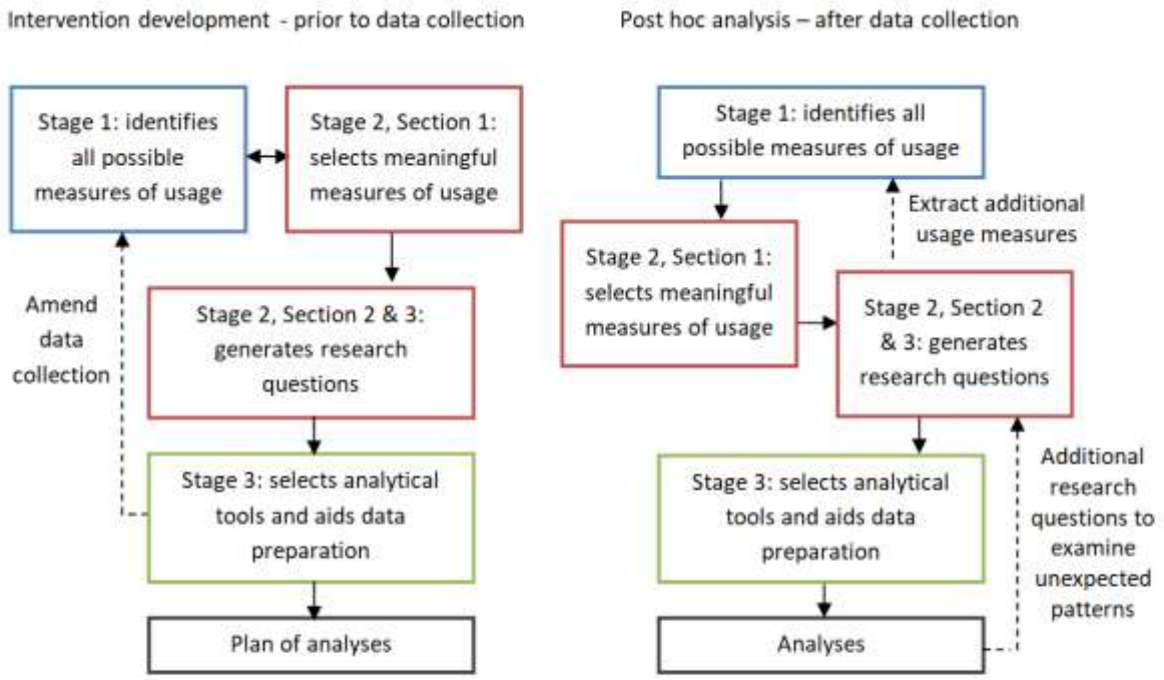


Figure 3. Stages of the AMUsED Framework

Dotted lines indicate optional paths to revisit if necessary

The framework focuses specifically on examining the relationships and associations between measures of usage and user characteristics, theoretical variables, behaviour and/or health-related outcomes. However, it is anticipated that analyses of usage would be considered in the context of a broader process evaluation that may examine how variables other than usage are associated with intervention outcomes (Moore et al., 2015).

3.4.2 Stage 1: Familiarisation with the data - identifying variables

Evaluation of a DBCI can produce large datasets containing information collected in a variety of formats. It may be necessary to collate relevant data across the datasets and compute new variables before usage analyses can be conducted. To simplify this process stage 1 proposes a checklist (see Section 3.7.1) comprising a set of generic questions that will support a comprehensive understanding of the structure, processes and content of the intervention in relation to data capture, contents of the datasheets, and factors related to trial implementation (e.g. participant recruitment, stage 1; 3) . When used during the development phase of an intervention the framework provides the opportunity to record and measure usage data that align with the proposed analysis plan. Ensuring efficient data capture at the outset can remove the need for extensive data cleaning and manipulation. When used for post-hoc analysis only, stage 1 can support the identification of appropriate usage variables and inform subsequent data cleaning and manipulation in preparation for analysis.

The usage data has been grouped into three categories. Intervention characteristics describes architecture, content, and expected workflow through the intervention, including intended usage (Mohr et al., 2014) (e.g. anticipated number of logins, number of available content components, number of pages within a tunneled section). Accrued data covers all data collected during the running of the intervention, such as logs of interactions (log-data, e.g. date and time of use, pages viewed and time spent on them) and user-entered data (e.g. self-report). Contextual data encompasses previous findings related to the intervention development and trial (e.g. factors affecting usage), and relevant external factors (e.g. national health promotion campaigns).

Accrued data is usually collected automatically and recorded in datasheets, making the variables easier to extract and analyse. Whilst some variables for intervention characteristics and contextual data may also be captured in this way (e.g. number of logins), it is anticipated that additional measures may be either conceptual or external, meaning that they would not automatically be recorded as analyzable measures. For example, a DBCI designed to increase physical activity may include pages addressing barriers to the target behaviour, such as not having enough time to exercise, or not having access to equipment. Pages containing advice to overcome these barriers may be distributed throughout the intervention, but are based on the same underlying theoretical concept (an intervention characteristic).Therefore, a new variable needs to be created capturing usage of these pages, and exposure to the theoretical-underpinning. By identifying these measures of usage in advance it is possible for additional data capture processes to be created, aiding efficient analyses after data collection. For example, dependent on delivery platform, code may be added to an intervention so that users who view these pages are recorded within a unique variable column in the log-data. Contextual data is also less likely to be automatically collected and recorded within the study. If a large-scale outbreak of a respiratory infection occurs during the trial of an intervention aimed at reducing transmission of such infections, one might want to assess the potential impact on usage data. As with intervention characteristics, it may be possible to add further measures in order to capture personal experience of the illness or impact in a broader context.

It should be noted that the range of data available to collect may differ depending on the software used to develop and/or deliver the DBCI. Using the framework during development to determine in advance which data is crucial may facilitate software development or else alternative work-arounds. For example, where software is unable to collate total time spent on selected intervention pages, it will be necessary to ensure time spent by page is readily available to collate this after data collection. Interventions will also vary greatly in architecture and structure dependent on design, software and delivery platform used (e.g. website, app, text-based). For example, a 'session' (stage 1; 1) may refer to a single login, a component of content available across multiple logins, or the amount of times a specific activity is accessed. The framework provides a structure broad enough to be applied to different interventions. However, it is not anticipated that all criteria in the checklists will apply to every DBCI. Where concepts and examples provided do not directly translate, researchers are encouraged to define them as relevant for their intervention and adapt the framework as needed. .

3.4.3 Stage 2: Selecting measures of usage and generating research questions for engagement

Establishing testable hypotheses is the pre-cursor to carrying out systematic analyses. The aim of stage 2 is to support the generation of specific research questions to drive hypotheses testing. Stage 2 is divided into 3 sections to reflect the increasing complexity of comprehensive usage analyses: section 1 helps define specific measures of usage (descriptive statistics); sections 2 and 3 generate research questions (bivariate and multivariate analyses).

3.4.3.1 Stage 2, Section 1: Descriptions of usage variables

The first section of stage 2 provides a non-exhaustive list of potential usage measures. Example questions on the checklist (see Section 3.7.2) demonstrate how measures of usage may be constructed.

The abundance of data and potential usage variables can encourage unsystematic data-dredging. Identifying and reviewing the range of usage measures available enables researchers to make informed and/or theoretically-driven decisions about what will be the most meaningful variables to include in any subsequent analysis plan. The process of familiarisation with intervention content and architecture (stage 1) may highlight considerations when selecting usage variables for analysis. For example, it may become apparent that certain sections of the DBCI were considered to be of greater importance during the planning process, such as components that are theoretically informed (e.g. pages containing BCTs/advice, goal-setting sections). Therefore, analysing the usage of these pages specifically would be more meaningful than analysing the total number of pages viewed. Alternatively, the intervention logic model may indicate that two theoretically-based components are considered to have equal importance, yet they may have differing amounts of content within them, meaning that users would spend more time on one than the other. On that basis, analysing the time spent on theory-based components may result in misleading conclusions about the impact of usage on health outcomes. In this scenario, a categorical usage metric may be more meaningful e.g. having completed or revisited the component. Analysing a single usage measure is unlikely to provide a comprehensive understanding of engagement for all users across an intervention. However, combining multiple usage measures in a systematic way will provide a more detailed understanding of how users engaged with the intervention and what patterns of usage are associated with intended outcomes.

During intervention development this process is undertaken prior to data collection and is therefore based on prospective data identified in stage 1 (bracketed numbers provided in stage 2, section 1 indicates their counterpart sections in stage 1). Given the considerable crossover, we anticipate that these sections will be completed iteratively (see Figure 3). The purpose of stage 1 is to identify all potential measures of usage available within the data, stage 2 then narrows that selection by considering which will provide the most informative understanding of usage for a specific intervention. The selection is informed by fundamental elements of the intervention highlighted in the planning process (Band et al., 2017; Yardley et al., 2015). For example, in a DBCI targeting weight loss, important measures of usage might be identified as entering weekly self-reports of weight, repeat use of recipe component, and time spent watching exercise videos. When carrying out post-hoc analyses, descriptive statistics for measures of usage identified in stage 2 may provide greater insight into which measures will be more informative (see Figure 3).

3.4.3.2 Stage 2, Section 2 & 3: Relationships between usage, participant characteristics, target behaviours and behavioural determinants

The remaining two sections of stage 2 (see Section 3.7.2) will guide the generation of specific research questions to assess how usage might be related to participant characteristics, behavioural determinants and target behaviour. Whilst this stage can be used to generate limitless questions to drive exploration of the data, the framework is instead intended to be used to help select the most important questions that will answer theory-driven hypotheses. Usage variables are considered in relation to participant characteristics (stage 2; 2), target behaviour and behavioural determinants (stage 2; 3), and behaviour change across the intervention (stage 2; 3). By answering these questions it is anticipated that patterns of usage which reflect effective engagement with a specific intervention can be described. The moderating effect of demographic, psychosocial and health factors (stage 2; 2) on the relationship between usage and outcomes are also considered in section 3 of stage 2. When defining these variables, it is intended that the framework be adapted to individual interventions, for example it is possible that a measure of usage (e.g. uploading ongoing health monitoring statistics) may also be the intended primary outcome (Band et al., 2017).

3.4.4 Stage 3: Preparation for analysis

The stage 3 checklist (see Section 3.7.3) supports the process of selecting appropriate types of analyses and analytical software, and the data preparation necessary to translate the research questions developed in stage 2 into a plan of analysis. Generic questions guide the researcher to consider broad issues such as available resources (stage 3; 1, e.g. timeframe, additional researcher support, analysis plan for efficacy), more specific issues of selecting appropriate type of analysis and analytical software (stage 3; 2), and data management (stage 3; 3, e.g. amalgamation, manipulation, cleaning).

Our experience suggests that traditional statistical methods are not always suitable for analysing the types of research questions generated by the framework. For example, while research is usually powered to analyse efficacy, it is frequently underpowered for the type of sub-group analyses needed for in-depth usage analyses. In addition, whereas analyses of amounts of usage (e.g. total time spent or number of logins) often lend themselves to traditional methods, examining patterns of usage (e.g. movement through pages) requires alternative methods to identify and inform subsequent statistical analysis. Therefore, techniques such as visualisation and process mining may be more informative as they can reveal patterns of usage within the data, such as workflow through an intervention, clustering by participant groups, and temporal details (van der Aalst, 2016.; Arden-Close et al., 2015; Morrison & Doherty, 2014; Morrison et al., 2018; Rahman et al., 2017). In applying stage 3 of the framework, the required data format for any analytical software should be considered.

3.5 Application of AMUsED: Two case studies

The following section provides researchers with practical examples of how the framework checklists can be applied in advance of or after data collection. The key findings and applications from utilizing the framework are highlighted below. The framework is necessarily comprehensive, and completion of the checklists creates a lot of data/information. Indeed, this is the very process by which it supports the generation of systematic and rigorous usage analyses. The completed checklists for both studies have been amalgamated to enable comparison, and are available as supplementary data (see Appendices B-D) so researchers using the framework in practice have detailed examples of its use.

3.5.1 Applying the framework during development: Germ Defence

PRIMIT (PRImary care trial of a website based Infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) was a large randomised controlled trial (RCT) that showed a DBCI to be effective at lowering the transmission of colds, influenza and stomach upsets within the home through increased handwashing (Little et al., 2015). The framework was used to inform and structure the process of updating and amending the intervention ready for dissemination as an open-access resource for use by the general public. As part of that process the intervention was renamed Germ Defence. The research team involved with the dissemination was already familiar with the intervention having worked on the design and evaluation of the PRIMIT study. (For full details of the PRIMIT intervention and evaluation trial please see Little et al., 2015).

3.5.1.1 Stage 1: Familiarisation with data

Applying stage 1 of the framework supported us to undertake a detailed review of the original version of Germ Defence (stage 1; 3.2) along with data collected from the prior RCT (stage 1; 2). This informed crucial updates to the collection of usage data and the generation of research questions, which we describe in the following sections.

Disseminating Germ Defence to the general public required us to strike a balance between obtaining informed consent to collect a minimal amount of data to support evaluation whilst still enabling easy access to key aspects of the intervention by users who may be less willing to engage with standard research procedures (stage 1; 3.1). Completing the stage 1 checklist also allowed us to identify: 1) how the intervention and consent procedures should be streamlined (stage 1; 3.1 & 3.2), and 2) what pertinent self-report data should be collected to enrich analyses of the automatically collected usage data and enable comparison with the prior RCT data (stage 1; 1.2).

3.5.1.2 Stage 2, Section 1: Descriptions of usage variables

We reviewed the range of possible usage variables and identified which would provide the most informative picture of how Germ Defence was accessed and used during dissemination. For example, the first component of the intervention contains compulsory tunnelled pages including a section for selecting handwashing goals. Examining dropout across this component and online consent pages, along with repeat use of goal setting section, will enable us to understand if/where users disengaged with the intervention. We then compared our list with the data collected from the prior RCT. This identified crucial amendments to the data capture process for Germ Defence that would otherwise have been missed. Specifically, data recorded on use of the goal-setting component was over-written when revisited, losing both user-entered data and our ability to view movement backwards and forwards through these pages (stage 1; 1.2). Identifying this issue in advance meant we were able to adapt the back-end processes to ensure the required data was captured.

3.5.1.3 Stage 2, Section 2 & 3: Relationships between usage, participant characteristics, target behaviours and behavioural determinants

Completion of stages 1 and 2 of the framework in parallel helped us to narrow our selection of usage-related questions to focus on behavioural determinants that were identified to be most strongly correlated with the target behaviour in the prior RCT. Since efficacy of Germ Defence has already been established from the prior RCT, the primary focus of the dissemination phase is to examine patterns of usage 'in the wild' and their relationship to baseline user characteristics. The following research questions are a selection of those generated (stage 2; 2):

- Which pages see the highest amount of dropout, including consent and baseline measures?
- How do users move through the goal setting pages and what goals do they select?
- Are baseline measures for handwashing, level of belief that handwashing will lower infection transmission and/or belief in ability to increase handwashing associated with usage?
- Do users' perceptions about the risk of infection to themselves or a household member relate to usage?
- Does the means through which users hear about the website relate to usage?

Self-report data on behavioural and psychological variables will be collected using an optional survey (stage 2; 3), in order to minimize potential drop-out. This could be subject to selection bias with significant differences in the characteristics of users choosing to complete or not complete the survey. Any analysis examining the association between usage and behavioural outcome/change in behavioural determinants will be undertaken with caution. However, accessing/completing the survey may be operationalized as a measure of usage (stage 2; 1), providing the opportunity to analyse relationships between intervention and survey usage (e.g. is viewing more intervention pages associated with completing the survey? stage 2; 2). A comparison of baseline characteristics will also enable a check of whether those who complete the follow-up survey are different to those who do not.

3.5.1.4 Stage 3: Preparation for analysis

The analytical tools available are SPSS for Windows (SPSS Inc., Chicago) and LifeGuide Visualisation Tool (LVT) (Arden-Close et al., 2015) (stage 3; 2, Appendix D). It is anticipated that there will be insufficient power for definitive hypothesis testing. Patterns of usage (e.g. repeat use, dropout across tunnelled pages) will be best explored using visual tools initially. As Germ Defence has been built using LifeGuide software the data produced will be compatible with the visualisation tool (stage 3; 3). Usage data collected from the intervention will need to be amalgamated and linked with self-report data from the optional survey. Thus, it was necessary to ensure that all users were allocated a unique non-identifiable numeric ID upon first access so that all data can be linked (stage 3; 3).

Using the research questions developed in stage 2 and considerations highlighted in stage 3, a full plan of analysis for Germ Defence was developed to inform efficient and systematic analysis post data collection. Applying the framework helped us prioritize research questions most relevant for the focus of the research (e.g. how interventions are accessed and used 'in the wild'), that would not be undermined by the constraints of using optional self-report measures.

3.5.2 Applying the framework for post-hoc analysis: Internet Dr

3.5.2.1 Overview

The framework was used to develop an analyses plan for usage data collected during a randomised controlled trial (RCT) of the Internet Dr, a DBCI to support the self-care of respiratory tract infections (RTIs) and reduce unnecessary GP visits. The RCT showed that users with access to Internet Dr were less likely to contact their GP about an RTI than those without access (Little et al., 2016). The usage analyses for Internet Dr will be conducted by researchers who were not involved in the original design, development and evaluation of the intervention. The framework enabled the researchers to understand the intervention and associated data collection, and construct systematic research questions to investigate usage. (For full details of the Internet Dr intervention and evaluation trial please see Little et al., 2016).

3.5.2.2 Stage 1: Familiarisation with data

Internet Dr is structured around three components of theoretically-based content. 'Doctor's Questions' and 'Common Questions' aim to support users who are unsure if their symptoms are serious and whether they are in need of medical treatment (stage 1; 1.2, 3.2) (Leventhal, Brissette & Leventhal, 2003). 'Treatment Options' is intended to increase self-efficacy for users who wish to manage symptoms they are finding distressing (stage 1; 1.2, 3.2) (Bandura, 1997). Applying stage 1 aided understanding of how these three components relate to the psychological theories underpinning the intervention and thus the proposed determinants of the target behaviour (illness perception, health locus of control, willingness to tolerate symptoms, treatment preferences, see Appendix B).

All content was available whenever users accessed the intervention across a 24 week period in winter (stage 1; 1.1). However, users were encouraged to login specifically during periods of illness to help manage their symptoms. Completing the checklist emphasized the importance of recognizing these two distinct purposes for accessing the intervention: 1) to view content whilst ill, 2) to view content when well perhaps out of curiosity. These differences in motivation to access the intervention when well/unwell may also be reflected in differences in patterns of usage.

3.5.2.3 Stage 2, Section 1: Descriptions of usage variables

Given the theoretically-based content of the three components within the intervention, usage of each was identified as relevant to understanding underlying mechanisms of action (e.g. number of users, number of pages viewed, time spent, number of re-visits) (Appendix C). For the 'Doctor's Questions' component, compulsory tunnelled pages are completed leading to illness management advice on the last page (stage 1; 1.1). Therefore, users of this component would, in theory, not benefit unless they had reached the final page, so completion and dropout were identified as important measures of usage for this component. It was also intended by design that users would view 'Doctor's Questions' first (stage 1; 1.2). Thus, analysing the order in which users visited the different components was important to understand whether the intervention was used as intended, and also how intended versus non-intended order of use was related to users perceptions of their RTI, their perceived ability to self-manage and whether they contacted their GP. As differences in users' motivations for accessing the intervention may lead to differences in usage patterns (stage 1), measures of usage identified in this component (e.g. number of pages viewed and time spent) should be described for three situations: usage when ill, usage when well, and across all usage.

3.5.2.4 Stage 2, Section 2 & 3: Relationships between usage, participant characteristics, target behaviours and behavioural determinants

Considerations from the previous sections helped form pertinent research questions for the remaining two sections of stage 2 (Appendix C). For example, relationships between viewing specific content and theoretical constructs/behavioural determinants will be examined. As motivations for use have been identified as potentially influential, associations between reasons for accessing the intervention and patterns of usage and/or personal characteristics will be explored (stage 2; 2). In addition, we will examine whether users followed the intended navigational paths and whether this was related to visiting their GP (stage 2; 3). Below are some example research questions from the usage analysis plan:

- Does usage of Doctor's Questions differ when intervention access is made when ill compared to when well? (e.g. starting the component, number of pages viewed, time spent, viewing the advice page)?
- Are baseline personal characteristics associated with intervention use when ill/not ill?

- Is viewing content during illness associated to lower GP visits?

Finally, we aim to identify whether viewing a specific piece or amount of content at a certain time (e.g. when ill) led to a user being less likely to contact their GP.

3.5.2.5 Stage 3: Preparation for analysis

SPSS for Windows (SPSS Inc., Chicago) and LifeGuide Visualisation Tool (LVT) (Arden-Close et al., 2015) will be used for data analysis (stage 3; 2, Appendix D). There will be sufficient power to analyse average usage of the intervention (e.g. by whole sample) and associations with behavioural determinants/target behaviour. Sub-group analyses are unlikely to be sufficiently powered (e.g. comparing usage/outcomes of users accessing content when ill versus when prompted by completion of interim study measures). Some of the identified patterns of usage include movement through the intervention (stage 3; 1) and will therefore be best explored visually (e.g. order in which the three content components were accessed). The datasheets are compatible with LVT. However, data is spread across several datasheets, requiring extraction, transformation and amalgamation prior to analysis (stage 3; 3).

Through completion of the 3 stages of the framework (see Appendices B-D), it was possible to breakdown a complex DBCI and develop a comprehensive usage analyses plan which will help identify what type of usage was successful in supporting self-management, for whom was it most beneficial, and at what time it was most influential.

3.5.3 Comparing case studies

Both interventions target behaviour associated with respiratory tract infections (RTIs). However, Germ Defence focuses on infection prevention and may be accessed at any time, whereas Internet Dr supports self-management of symptoms whilst infected with an RTI. Despite the differences in their architecture, content and function, the framework was suitably generic to be applied to both interventions. Although the same stage 1 checklist was applied to both interventions, it enabled two completely different processes: for Germ Defence stage 1 helped shape structural changes to the intervention and data capture processes; for Internet Dr, it enabled understanding of a previously unfamiliar and complex intervention and the accompanying datasheets (Appendix B). Despite these different requirements, the checklist was comprehensive enough to fulfil both needs and lead to greater insights, such as realizing that a key component of Germ Defence (the goal-setting component) was not capturing data as required, and understanding that Internet Dr was designed to be used during illness but could be accessed at any time. Completing this first stage was the most complex and time consuming of the three stages for both interventions. However, through the thorough understanding of the intervention gained from stage 1, the following stages were easier to complete as the information was readily available to fit the generic questions. For example, having identified the theoretical underpinning of the 3 components of Internet Dr, their related measures of behavioural determinants and expected relationship to GP contact, generating research questions to examine how usage related to changes in behaviour and behavioural determinants was both simple and quick.

Through completing the checklist for stage 2, section 1 it is apparent that operationalizing usage in terms of amount was valid for both interventions (e.g. number of logins, number of pages viewed, time spent on them, Appendix C). This suggests that describing the extent to which an intervention has been used is a necessary first step for examining usage, and that the number of pages viewed and time spent on them may inform our understanding of different styles of engagement. For example, spending more time on or revisiting pages may be indicative of higher levels of interaction compared to viewing pages briefly. However, as previously discussed focusing on broad-based, summative descriptions of usage alone may not be sufficient to understand how the intervention supported change in target behaviours/outcomes. For example, distinguishing and comparing usage of the three theoretically-based content components of Internet Dr will aid understanding of the potential mechanisms of action within the intervention.

The differences in research questions generated from stage 2 highlight the differences in structure between the two interventions. Germ Defence is a standalone intervention requiring access only once, with an optional follow-up survey. Once completed, the data generated from using the intervention will provide a snapshot of behaviour at that time. On that basis, research questions focus on user characteristics and behavioural determinates/ target behaviour at baseline (Appendix C). In contrast, in addition to self-report and log-data over a six month period, Internet Dr users' GP notes providing information for the year before and after the trial commenced were also collected. This depth and length of duration of data enables different research questions including consideration of behaviour prior to the trial and for some months after.

Stage 2 of the framework highlighted the relevance of examining patterns of usage (e.g. movement through the intervention) and sub-groups usage analyses. As both interventions have insufficient power to analyse sub-group usage, and patterns of usage lend themselves to visual exploration (Appendix C), this supports the use of contemporary visualisation tools in addition to traditional statistical methods.

Both interventions include research questions examining relationships between participant characteristics and usage, and whether any of those characteristics moderate the relationship between usage and target behaviour. Through comparison of these similar analyses across multiple interventions it will become possible to build up a pattern of how personal characteristics may influence DBCI usage, leading to generic learning points to inform future intervention design. This may also be the case for usage analyses of interventions with similar aims (e.g. self-management of illness) or similar theoretical underpinning, behavioural determinants or BCTs. Once a body of research is assembled, it would be possible to use the framework in advance to structure data capture and analysis so that it is comparable with prior interventions and published research.

3.6 Discussion

The AMUsED framework aims to support detailed and systematic analysis of DBCI usage. The framework comprises three stages of checklists for researchers to: 1) understand the intervention’s design, theoretical-underpinning and data collection processes, 2) define meaningful variables to assess usage, and generate both broad and fine-grained research questions to examine relationships between usage, participant characteristics, and target behaviour/behavioural determinants, 3) prepare datasheets and consider appropriate software for analysis.

The framework has been applied to two DBCIs: Germ Defence promotes RTI prevention, and Internet Dr supports self-management of RTI symptoms. Using the framework whilst preparing Germ Defence for public dissemination identified necessary amendments to data capture processes. For Internet Dr, the framework helped guide a research team who were previously unfamiliar with the intervention design and data to devise a comprehensive usage analysis plan. The case studies demonstrate the flexibility of the framework to be applied to different interventions, and the advantages of using the framework both before and after data collection.

3.6.1 Implications

The AMUsED framework checklists provide researchers with easily applied templates for carrying out detailed usage analyses of DBCIs. The framework supports the level of rigor in reporting DBCI content and findings called for by current guidelines from the MRC, CONSORT-EHEALTH and TIDieR (Craig et al., 2008; Eysenbach & CONSORT-EHEALTH Group, 2011; Hoffmann et al., 2014; Moore et al., 2015). The checklists extend upon and draw together previous work on categorizing digital intervention usage data and selecting research questions (Danaher et al., 2006; Eysenbach, 2005; Paz Castro et al., 2017; Sieverink, Kelders, Poel, et al., 2017; Taki et al., 2017) by providing a systematic and comprehensive process for researchers to follow. The process can be incorporated into existing DBCI development methods such as the Person-based approach (Yardley et al., 2015) and the BIT model (Mohr et al., 2014), enabling pre-testing of data capture processes to support theory-based hypotheses testing.

The framework encourages usage analyses that will broaden our understanding of mechanisms of action underlying a specific DBCI, explaining the relationships between user characteristics, patterns of usage and behaviour change. Through this process it may be possible to identify effective engagement, finding the level of usage necessary for a specific intervention in order to change the target behaviour (Ainsworth et al., 2016). This will lead to DBCIs being developed to be more concise, efficient and targeted, making them less arduous for the user and supporting higher rates of uptake and engagement.

3.6.2 Limitations

The framework has been developed and tested using web-based interventions built using the same software (Williams, Yardley, Weal, & Willis, 2010) which captures extensive log-data and has the ability for researchers to write additional code in order to capture tailor-made usage measures. DBCIs may be delivered across a wide variety of platforms (e.g. text messaging, apps and websites) and developed using different software. This leads to substantial variation in design, the manner in which they are written or coded, and the availability and format of data collected. The framework is flexible enough to be applied across diverse interventions, and sufficiently detailed to generate specific testable hypotheses for most DBCIs. However, we welcome other researchers to use the framework and build upon it based on their experience.

The AMUsED framework focuses on the analysis of measures of usage, as one facet of engagement. Where objective measures of physiological reactions (e.g. cardiac activity, eye tracking), or subjective self-report measures of engagement are available, it is hoped that future research may examine these alongside usage data and develop the framework further to incorporate them, thereby increasing our ability to explain not just the role of usage, but engagement more broadly (Perski et al., 2017). It is also our hope that the framework will be applied as part of a mixed-methods approach, triangulating usage analyses with insights and experiences collected qualitatively (Michie et al., 2017; Perski et al., 2017; Yardley et al., 2016).

The framework has been applied to ensure adequate data collection when used during the development phase of DBCI. However, the framework also has the potential to be used to inform study design to answer empirical questions on effective engagement (e.g. Multiphase

Optimization Strategy MOST, Sequential Multiple Assignment Randomized Trial SMART (Collins, Murphy, & Strecher, 2007). Although the current paper does not address this application, it provides an avenue for future research for the wider application of the framework.

3.6.3 Conclusions

The AMUsED framework offers a systematic process for carrying out in-depth usage analyses. The aim of the framework is to capture and formalize the techniques used by experienced researchers to support researchers who are new to conducting usage analyses, or new to a particular intervention, in deciding how to assess usage data that will be or has been collected. Using the framework will benefit researchers by lowering the possibility of overlooking key questions and making the reporting of usage analyses more efficient, leading to a quicker turnaround for publishing. The checklists provide the means to increase transparency and make findings easier to replicate, whilst discouraging unsystematic data dredging. The process will also encourage greater detail and consistency in the reporting of usage and engagement, making it easier to apply the findings to a wider context (Nelson et al., 2016) and enabling comparison across different interventions and evaluation studies. The framework helps to operationalize and measure usage in ways that will better inform our understanding of engagement with a DBCI, encompassing broad measures of usage by the whole sample through to specific theory-based usage variables and usage by sub-groups based on personal characteristics. It guides insight into which components of an intervention worked, and how they interacted with users' personal characteristics. Finally, by using the framework it may be possible to identify the extent of usage required to support changes in behaviour/health-related outcomes, and thus an understanding of what constitutes an effective level of engagement for specific interventions.

3.7 The AMUsED framework checklists

- Stage 1: familiarisation with the data
- Stage 2: selecting usage variables and generating research questions
- Stage 3: preparation for analysis

3.7.1 Stage 1 checklist

Familiarisation with the data – identifying variables	
Generic questions by data type	Intervention Name:
1. Intervention characteristics. Data for intervention architecture and content.	
1.1. Workflow. Intervention structure and expected participant interaction and navigation through the intervention.	
How many logins/sessions are available?	
When are they available?	
Are new sessions released depending on time elapsed or task-completion?	
Are there limitations on the availability of the intervention?	
Is the purpose of a session to collect self-report measures and/or use the intervention?	
When is the intervention considered to be finished?	
What prompts are used to encourage usage (e.g. emails, texts, notifications) and when are they sent?	
Does the intervention contain ‘tunneled’ (compulsory) sequences of pages which users have to view to move forward?	
Are users able to select linked components they wish to view, and avoid others?	

1.2. Content. Content available within the pages of the intervention.

What are the components available?	
What is the aim of each component and are they based on underlying theoretical constructs?	
In what order is it anticipated the components will be used?	
What interactive features are available (e.g. forums, videos, printable information)? How long should they take to complete?	
Are all components/features available to all users throughout the intervention or are some tailored for specific times or users?	
Which pages are for collecting self-report measures or for administrative purposes (e.g. questionnaires, login, password change)?	
Are there specific pages to mark the start and end of sessions?	
Which pages contain BCTs (e.g. information, planning, feedback) and what are they?	
In which sessions are they available?	
Can specific BCTs be identified on particular pages or groups of pages? How many groups are there?	
Do any of the pages have response options to collect information in addition to baseline/follow-up measures? What data is collected?	

2. Accrued data. Data collected during an intervention.

2.1. Self-report. Users' self-reported responses collected across various stages of the trial.

When are self-report questionnaires collected (e.g. weekly logins, monthly symptom information, follow-up at 6 months)?

What demographic information is available (e.g. age, gender, education)?

Which measures are specifically related to the target behavior and how often are they collected?

Which measures of beliefs influential on the target behavior are collected and when?

Are measures of health collected (e.g. conditions which may impact on target behavior or are co-morbid) and psychosocial factors (e.g. anxiety, illness perception, motivation)?

Are additional measures collected at follow-up (e.g. satisfaction, adherence)?

2.2. Log-data. Information automatically collected through engagement with an intervention.

What data is the software platform able to record?

Are number, date and time of logins available by individual user?

Are individuals' total durations of usage accessible?

Are the number and time of usage prompts recorded?

Are there details for which pages were viewed, the sequential order and time spent viewing?

2.3. External data. Data collected independently but alongside intervention usage.

How and where is the data collected (e.g. GP or support staff notes, lab reports, other digital data such as activity or location trackers)?

What data is collected?

Which of these measures relate to or may impact on the target behavior?

3. Contextual data. Data indirectly related to the running of the intervention which may be influential over usage and analysis.

3.1. External factors. Structures and events which may influence participation in the intervention.

How are users recruited to the intervention?

Did any specific large-scale events, with the potential to impact on the intervention, occur during the period of the intervention (e.g. changes in treatment, health campaigns, illness outbreak, technical issues with the intervention)?

3.2. Previous theory and findings. Results of behavioral analyses carried out during intervention development (e.g. logic models), and analyses of clinical outcomes if available.

What are the hypothesized mechanisms of the intervention (e.g. as specified in the intervention’s logic model)?

Which factors are identified as important in qualitative research, and can they be related to the variables collected in the trial (e.g. preferences for specific pages)?	
Which variables are identified as relating to outcomes (e.g. behavioral determinants, theoretical constructs, health factors)?	

3.7.2 Stage 2 checklist

Selecting usage variables and generating research questions	
Generic questions	Intervention Name:
1. Descriptions of usage variables. Which usage variables are relevant to the intervention and in which format (e.g. number of users/sessions, duration, percentage of total, dichotomous)?	
Completing intervention/trial period (stage1; 1.1 & 2.2). E.g. How many users complete the trial? What is the average time taken to complete?	
Logins or sessions where the intervention was accessed (stage 1; 1.1 & 2.2). E.g. How many users start/complete each login/session? How long does it take to complete each session? How many pages are viewed within the session? Which session has the highest proportion of pages viewed, or duration of time spent on it?	
Date of login and usage. E.g. When do users login? What time of year? Are there changes in frequency of logins?	
Time of day of login and usage. E.g. What time of day is usage? Are users more likely to spend longer on the intervention at certain times?	
Days/weeks of usage (stage1; 1.1 & 2.2). E.g. For how many days/weeks out of the total is the intervention accessed for? How many times within a week is the intervention accessed? Are there repeated uses within the same day?	

Response to prompts/notifications (e.g. requests to login, email, text, upload data) (stage1; 1.1 & 2.2). E.g. How many responses are sent? How long after receiving notification do users take to log-in or respond?	
Features/linked menu components used (stage 1; 1.1, 1.2 & 2.2). E.g. How many features/components are accessed? How many users access each one? Which are completed and by how many users? Which feature/component has the highest proportion of pages viewed or time spent? What order are they viewed in? Is this the anticipated order? Which have the highest proportion of drop-out?	
Revisiting components/features (stage 1; 1.1, 1.2 & 2.2). E.g. Are any used repeatedly? How many times are they revisited, and for how long? Which are most revisited?	
Type of content/BCTs used (excluding administration pages) (stage 1; 1.2 & 2.2). E.g. How many groups of pages with similar content are accessed and by how many users? How many pages within the group are used? How many users view each page? Which groups of pages have the highest proportion of views? Which pages are viewed at each login, and when is the largest amount of pages viewed? Which pages have higher drop-out?	
Completing ongoing measures (e.g. monthly questionnaires, response options within content pages, uploading information or text responses) (stage 1; 2.1 & 2.2). E.g. How many users complete ongoing measures? When do they complete them? Do they also access the intervention at that time?	
External device usage (e.g. wearables and other sensor technologies) (stage 1; 2.3). E.g. How much time is spent with the device? How many times is it used? What number of	

days/weeks is it used for?	
2. Relationships between usage and participant characteristics. Are users' demographic, physical or psychosocial characteristics at baseline related to intervention usage?	
Are any characteristics at baseline related to usage? E.g. Is anxiety associated with revisiting features? Is current health related to usage of external devices? Are users who spend more time on the intervention older than those who spend less time? Which characteristics are associated with drop-out?	
Are any contextual factors associated with usage (stage1; 3)? E.g. Is manner of recruitment related to usage?	
Do high/low users differ by other usage factors? E.g. Do users who spend more time on the intervention view more types of content than users who spend less time? Is usage of an external device related to intervention usage?	
3. Relationships between usage, behavioral determinants, and target behaviors. Which usage variables are associated with follow-up measures for target behavior and behavioral determinants? Which usage variables help explain changes in behavior across the intervention?	
Are baseline measures for behavioral determinants/target behavior related to usage? E.g. Is the number of days the intervention is used for related to a behavioral determinant? Do users with low target behavior spend less time on the intervention?	
Which usage variables are related to behavioral determinants/target behaviors and at follow-up? E.g. Do users who view a group of pages containing a specific BCT score higher/lower for the associated behavioral determinant? Is completing/not completing a particular component associated with target behavior at follow-up? Is the time spent on a	

session related to target behavior?	
Is usage associated with measures for acceptability/satisfaction at follow-up? E.g. Are high levels of satisfaction associated with accessing more pages? Do users with low satisfaction spend less time using external devices?	
Do users who report positive changes in behavioral determinants/target behavior from baseline to follow-up use the intervention differently to those who do not? E.g. Do users who report positive increases in a behavioral determinant view more pages from a specific component containing an associated BCT? Do users who report positive behavior change spend more time on the intervention?	
Are relationships between usage and target behavior moderated by demographic, psychosocial or health factors? E.g. Is the relationship between time spent on the intervention and target behavior altered when moderated by anxiety?	
What level of usage is necessary for ‘effective engagement’? E.g. Do outcome measures plateau after viewing certain content, or after a certain amount of time or sessions completed?	

3.7.3 Stage 3 checklist

Preparation for analysis	
Generic questions	Intervention Name:
1. Resources	
What is the timeframe for completing the analyses?	
What resources are needed? E.g. additional research time, expertise	
Is a plan of analysis already available? How does the analysis plan developed using the framework compare to that plan? Are changes or updates needed?	
Is ethical clearance in place to carry out usage analyses?	
2. Selecting types of analysis and analytical software	
Will the usage data be triangulated with qualitative data?	
What analytical tools are available?	
Is there sufficient statistical power to answer the planned research questions?	
Can the selected measures of usage be analyzed using the available tools? Is bespoke software necessary (e.g. visualisation techniques)?	
3. Data preparation	
When is the data available?	

Is the data raw or has it been used/cleaned previously?

How many datasheets are there? Will these need to be amalgamated?

Is the data structured to work with the tools available? What formats are the datasheets in (e.g. excel, .csv) and will they need converting for analysis?

What preparation does the data need (e.g. cleaning, anonymizing)?

Are all variables readily available or will they need extracting/transforming/recoding?

Is the data in the right format to answer the research questions? Will it need adapting (e.g. continuous variables changed to categorical)?

Chapter 4 Paper 2: How digital interventions support self-care of minor ailments: A process evaluation of the Internet Dr intervention for respiratory tract infections.

Miller, S., Yardley, L., Smith, P., Weal, M., Milton, A., Stuart, B., Little, P., & Morrison, L. (2020). How digital interventions support self-care of minor ailments: A process evaluation of the Internet Dr intervention for respiratory tract infections. *Journal of Medical Internet Research Preprints,* 24239. https://doi.org/10.2196/preprints.24239

Currently in 2nd round of review with Journal of Medical Internet Research.

4.1 Abstract

Background: Around 57 million doctor appointments annually in the UK are for minor ailments that could be self-cared for by patients. As well as taking up healthcare resources, patients experience increased anxiety, lowered confidence and inconvenience. The ‘Internet Dr’ is a digital intervention developed to support patients to self-care for respiratory tract infections. In a randomised controlled trial recruiting patients registered with General Practices, those with access to the intervention had fewer visits to their doctor for respiratory tract infections. Having established intervention efficacy, further examination of the data collected in the trial is required to understand how the intervention was successful.

Objective: This paper reports a process evaluation of usage of the ‘Internet Dr’ by the intervention group. The evaluation demonstrates how meaningful usage metrics (ie, types of interaction that are specific and relevant to the intervention) can be derived from the theoretical principles underlying the intervention. These metrics are used to examine whether these interactions were effective in supporting self-care for respiratory tract infections, for whom and at what time.

Methods: The ‘Internet Dr’ trial recorded patients’ characteristics and usage data over 24 weeks. At follow-up users reported whether their levels of enablement to cope with their illness changed over the trial period. Medical Research Council process evaluation guidance and AMUsED framework checklists were applied to structure research questions to examine associations between usage and enablement.

Results: Viewing pages containing advice on caring for respiratory tract infections was identified as a meaningful metric for measuring usage of the intervention. Almost half the users (*n*=616, 42.32%) viewed at least one advice page, with most people (*n*=478, 77.60%) accessing them when they initially enroled in the study. Users who viewed an advice page reported increased enablement to cope with their illness as a result of having participated in the study compared to users who did not view advice pages (*M*=2.12 vs *M*=1.65, *MD* =.47, 95% CI [.08, .86]). Users who had visited their GP for a respiratory tract infection in the year prior to the trial were a target population, and analyses revealed that this group were more likely to access advice pages (odds ratio 1.349, 95% CI 1.159, 1.571, *P*<.001).

Conclusions: The process evaluation identifies viewing advice pages as associated with increased enablement to self-care, even when accessed in the absence of a respiratory tract infection, meaning that dissemination activities need not be restricted to targeting users who are ill. The intervention was effective at reaching the target population of users who consulted their GP previously. However, attrition prior to advice pages was high, highlighting the necessity of prioritizing access during the design phase. These findings provide guidance on how the intervention may be improved and disseminated, and have wider implications for minor ailment interventions.

Trial registration: ISRCTN91518452

4.2 Introduction

Minor ailments are defined as non-serious health conditions that may be cared for by patients (e.g. back pain, respiratory tract infections [RTIs], headache, stomach upsets) (Banks, 2010). However, an estimated 57 million unnecessary visits to General Practitioners (GPs) in the National Health Service (NHS) occur every year in the UK as patients seek advice for managing these conditions (Proprietory Association of Great Britain, 2016). The strain this places on primary care resources is well documented (Nazareth & Murray, 2010; Richardson et al., 2018), but there is also a cost to the patient through increased anxiety, lowered confidence and inconvenience (Banks, 2010). Promoting self-care for these ailments would help alleviate the stress on both primary care and patients, by helping patients to understand and feel more enabled to cope with their health (Nazareth & Murray, 2010).

Many patients already use online resources for guidance with health issues (Tan & Goonawardene, 2017). However, credible, evidence-based interventions are needed to ensure potentially serious infections are identified and users advised to consult a health care professional (HCP) when necessary. Interventions aimed solely at increasing users’ knowledge about their illness have shown only limited effects on increasing self-care (Richardson et al., 2018). Instead, calls have been made for interventions that address barriers to self-care, such as patients feeling distressed about their symptoms and not knowing how to treat them (Richardson et al., 2018; Yardley et al., 2010). Theoretically-based digital health interventions have the potential to address these barriers, and offer the advantages of providing ongoing support at a time and place that is convenient to the user. Internet Dr is a digital intervention to support appropriate self-management of RTIs (Little et al., 2016). The intervention content is theoretically-underpinned and contains tailored advice on self-caring for RTI symptoms, as well as a symptom checker to identify serious illnesses including meningitis and sepsis (Yardley et al., 2010) (see Intervention section in Methods for more details). Content was designed to address previously identified barriers to self-care: 1) uncertainty about the need for medical treatment; 2) distress caused by the symptoms (Richardson et al., 2018; Yardley et al., 2010).

A Randomised Controlled Trial (RCT) of the Internet Dr was carried out over the winters of 2012 and 2013, with 3044 participants recruited randomly from lists of all patients registered at a selection of General Practices in southern England. Users who had access to the intervention had fewer GP consultations for an RTI compared to those in the control group (n=239, 15.2% vs n=304, 18.3%, multivariate risk ratio 0.71, 95% CI 0.52 to 0.98, p=0.048), despite both groups having equivalent occurrence of illnesses (Little et al., 2016). This means that more users in the intervention group decided to self-care for their symptoms. As well as GP visits, self-report scores for the Patient Enablement Index (PEI) were also collected as an outcome measure at follow-up to capture the psychological benefits for patients of using the intervention (Howie, Heaney, Maxwell, & Walker, 1998). The PEI items asked users to reflect on perceived changes that occurred as a result of having participated in the study: e.g. thinking about the kinds of symptoms we have asked about in this study, compared with before you took part in this study, do you feel you are able to help yourself: same or less; better; much better? Having previously focused on the intervention's impact on health service usage (Little et al., 2016), a process evaluation of the RCT data is required to understand the psychological changes and behavioural engagement with theoretically-underpinned content of the intervention which led to the intervention group's increased ability to self-care.

Process evaluations aim to provide insight in to what parts of an intervention work, for whom and under what conditions (Bonell, Fletcher, Morton, Lorenc, & Moore, 2012; Craig et al., 2008). This may be achieved by examining the underlying intervention mechanisms that are anticipated to lead to positive outcomes and the impact of context on the implementation of an intervention (Bonell et al., 2012; Craig et al., 2008). Logic models are often used to map the intervention content, theoretical underpinning, anticipated mechanisms of action and outcomes (Fletcher et al., 2016), thereby identifying core research questions or hypotheses to address within a process analysis (Moore et al., 2015). By explaining the mechanisms and effects of context, process evaluations have the potential to inform future intervention development and dissemination, and advance our understanding of intervention theory (Bonell et al., 2012; Craig et al., 2008).

Quantitative usage data collected automatically during interaction with a digital intervention (ie, log-data) has the ability to provide a rich source of metrics for usage analyses (Miller et al., 2019; Pham, Graham et al., 2019; Sieverink, Kelders, Poel, et al., 2017). Although widely used, broad, summative measures of usage, such as time spent or number of pages/components viewed in an intervention, have been criticized for their inability to explain how usage led to positive outcomes (Ainsworth et al., 2016; Enrique, Palacios, Ryan, & Richards, 2019; Miller et al., 2019; Yardley et al., 2016). In addition, the breadth of potential usage metrics available means that there is a danger that inferential analyses that examine all of these variables will produce results that do not relate meaningfully to the theoretically designed intervention architecture, and are therefore unable to offer specific practical and actionable recommendations to optimize future intervention design (Yardley et al., 2016). A clear rationale for choosing usage metrics is necessary to understand what is being measured and can be inferred from analyses (O'Brien & Toms, 2008; Sieverink, Kelders, Poel, et al., 2017). On that basis, arguments have been made for identifying usage metrics that are meaningful to the intervention, rather than data dredging (Miller et al., 2019). This means determining types or patterns of usage that are specific to an intervention's structure and the target behaviour, and are able to examine usage of theory-based content (Miller et al., 2019; Yardley et al., 2016). For example, by isolating usage of a specific component or set of pages aimed at improving users' self-efficacy for carrying out a target behaviour, it is possible to examine the relationships between having viewed that component, reported changes in self-efficacy, and behavioural outcomes.

The framework for Analysing and Measuring Usage and Engagement Data (AMUsED) (Miller et al., 2019) was developed to support systematic usage analyses of digital interventions by guiding researchers through three stages of planning and carrying out analyses. Stage 1 focuses on familiarisation with the intervention architecture including content, structure and data collection. Through a list of generic questions in the first section of Stage 2, researchers identify available metrics with which to measure usage, covering both summative measures (e.g. number of times the intervention was accessed, completing the intervention, amount of time spent) and more in-depth measures (e.g. type, frequency and completion of theoretically-based content). Researchers are then encouraged to consider these variables alongside the information from Stage 1 and to identify usage metrics that are relevant to the intervention structure, theory-based content and target behaviour, and that are most likely to provide insight in to how the intervention was effective, may be improved and implemented. Sections 2 and 3 of Stage 2 then address how these metrics may be utilised in inferential analyses with self-report measures for users characteristics and the target behaviour/outcomes. Stage 3 focuses on planning compatible data collection to ensure that analysis using appropriate analytical software is both possible and less onerous. The framework has previously been used to shape data collection for other digital interventions (Miller et al., 2019), but this process evaluation reports the first application to a usage analysis.

In line with Medical Research Council (MRC) guidance (Moore et al., 2015), this paper reports the process evaluation of the Internet Dr RCT, including a detailed usage analysis structured by the AMUsED framework (Miller et al., 2019). The aims of the evaluation are: 1) to identify measures of usage that are meaningful to the intervention; 2) to examine which parts of the intervention worked, for whom and at what time. These results will generate guidance on how the design, implementation and dissemination of the Internet Dr can be improved, but will also have generic implications for guiding the successful design and development of other digital interventions promoting self-care for minor ailments.

4.3 Methods

4.3.1 Internet Dr trial design

An open pragmatic parallel group RCT of the Internet Dr digital intervention had been carried out previously (Paul Little et al., 2016). After completing online enrolment to the study and baseline measures, participants were randomised using computer-generated random numbers to either the intervention group who had access to the website, or the control group who did not. Having completed outcome measures at 24 weeks, participants in the control group were able to view the intervention. For full details of the Internet Dr RCT and findings from the primary analysis see Little et al. (2016). Trial registration number: ISRCTN91518452, ethics approval from South West Medical Research Ethics Committee, UK Health Departments' Research Ethics Service.

4.3.2 Participants

Adults (aged ≥18) registered with general practitioners within NHS Primary Care were recruited to the RCT by postal invitation. Patients with severe mental health problems or terminal illness were excluded. Participants needed to have access to the internet, with only one participant per household taking part. The process evaluation examines only participants who were randomly allocated to the intervention group and therefore had access to the intervention during the 24 week trial period.

4.3.3 Process evaluation design

A plan for carrying out a complete process evaluation of data collected during the Internet Dr RCT was designed and conducted in line with MRC guidance (Moore et al., 2015), using the AMUsED Framework (Miller et al., 2019). Based on the AMUsED framework checklists, the intervention's structure, theoretical underpinning and data collection points were collated (Stage 1, Appendix E). All available usage metrics were considered in relation to the information in Stage 1 to ascertain which types of usage would be most meaningful to the intervention (Stage 2, Section 1, Appendix F). A comprehensive list of research questions was then generated to examine associations between the meaningful measures of usage, user characteristics and outcomes (Stage 2, Sections 2 and 3, Appendix F). The questions were refined based on the logic model (see Intervention section). The most appropriate analytical tools for examining the research questions were selected, and necessary data preparation considered (Stage 3, Appendix G).

The process evaluation team combined expertise in psychology, primary care, statistical analyses and computing. Four team members were previously unfamiliar with the intervention. The other four researchers had been involved in various stages of the Internet Dr development and primary outcome evaluation of the trial (Little et al., 2016), and advised on intervention content, logic model, data capture processes and analyses. The first author had previously developed the AMUsED framework for application in process evaluations but was not familiar with the Internet Dr intervention.

4.3.4 Measures

Participants were requested to complete online: baseline measures at the start of the trial; interim questionnaires every 4 weeks on RTI occurrence; outcome measures at 24 weeks. Actual GP visits prior to and during the trial were collected after 1 year from participants' GP records. Log-data for individual users and sessions were collected during the trial (e.g. pages accessed, time spent, order of pages viewed).

Modifiable psychological characteristics thought to underlie decisions to self-care were measured at baseline and follow up to capture any changes over the trial period that may help to explain outcomes (Theory of Planned Behaviour, TPB; Ajzen, 1991); attitudes; norms; perceived behavioural control (PBC), and beliefs in the necessity of HCPs; Health Locus of Control (Wallston, Strudler Wallston, & DeVellis, 1978); Krantz Health Opinion Survey (Krantz et al., 1980)). Trait anxiety (Health Anxiety Inventory; Salkovskis, Rimes, Warwick, & Clark, 2002) and intentions to use and follow intervention advice (TPB; Ajzen, 1991) were measured at baseline. Experiences of accessing and using the intervention were collected at follow-up (Problematic Experiences of Therapy Scale (PETS; Kirby, Donovan-Hall, & Yardley, 2014), along with the psychological outcome measure of how much users felt their ability to cope with an RTI had changed over the course of the trial (Patient Enablement Index, PEI; Howie et al., 1998). Full details of psychological measures and response items are available in Appendix H.

4.3.5 Internet Dr Intervention

Internet Dr is a web-based digital intervention developed using LifeGuide Software (Williams et al., 2010). All participants were encouraged to log-in as soon as they received the invitation letter to the study from their GP. Having completed trial enrolment and baseline measures, participants allocated to the intervention group were able to access the whole intervention immediately and at any point throughout the study. In addition to completing interim questionnaires, users were encouraged to log in again should they experience an RTI. The intervention was developed between 2008-2009, prior to the widespread use of smart phones.

4.3.5.1 Intervention content

Internet Dr comprises of three theory-based components offering varying levels of tailored advice (see Figures 4 & 5). For full details and examples of content see Yardley et al. (2010). The ‘Doctor’s Questions’ component contains a symptom checker with detailed questions about users’ symptoms. Based on these answers, users are shown one of three tailored advice messages: 1) ‘Your symptoms could be a sign of a serious condition that needs urgent care, ring NHS Direct immediately’; 2) ‘You should contact NHS Direct for further advice’; 3) details on how to self-manage symptoms with a recommendation to re-visit the website should their symptoms not improve or deteriorate further. NHS Direct was a triage phone service where patients were advised whether they need to visit a hospital or their GP for their symptoms, this has since been replaced by NHS 111. Where patients are recommended by NHS Direct to contact their GP, this information is not automatically transferred to the patient’s GP notes. This component of the intervention ensures that users with potentially serious infections receive the required treatment. The ‘Common Questions’ component provides answers to ten frequently asked questions about RTIs (e.g., how can I tell if my symptoms are due to a cold or flu?). Questions of interest are chosen by the user but there is no tailoring in the answers provided. Both components are informed by Leventhal’s Common Sense Model of Self-regulation of health and illness (Leventhal et al., 2003), and aim to support users who are unsure if their symptoms are serious and whether they need medical treatment.

The ‘Treatment Options’ component supports users to manage any distress they have about their symptoms. The content is informed by Bandura’s Social Cognitive Theory (1997) to increase users self-efficacy for managing their symptoms independently. This section offers tailored advice on self-managing an RTI dependent on the symptom selected and preferred type of treatment (ie, without medication, medication from pharmacy, boosting the immune system). Whilst this component offers advice based on the type of treatment selected by the user, it is less tailored than Doctor’s Questions where the advice is specific to the individual and their need to consult the NHS.

Although each component is structured around a psychological theory and aimed at a specific barrier to self-care, the advice pages draw upon common behaviour change techniques (Michie et al., 2013). These include: instructions on how to perform the behaviour; information about health consequences; regulation through pharmacological support and by reducing negative emotions.

4.3.5.2 Intervention structure

All three intervention components are designed so that users are required to navigate through specific pages prior to reaching the RTI management advice (see Figure 4). However, each component is structured differently with varying numbers of prerequisite pages, therefore requiring differing levels of effort to access advice. For example, due to the high level of tailoring required for the symptom checker, a user may view up to ten pages of questions within Doctor’s Questions before reaching an advice page. In contrast, users accessing Common Questions need only view one page before accessing advice. The advice pages are designed to be stand-alone, such that viewing a single page from any component may provide the user with the necessary support to self-care.

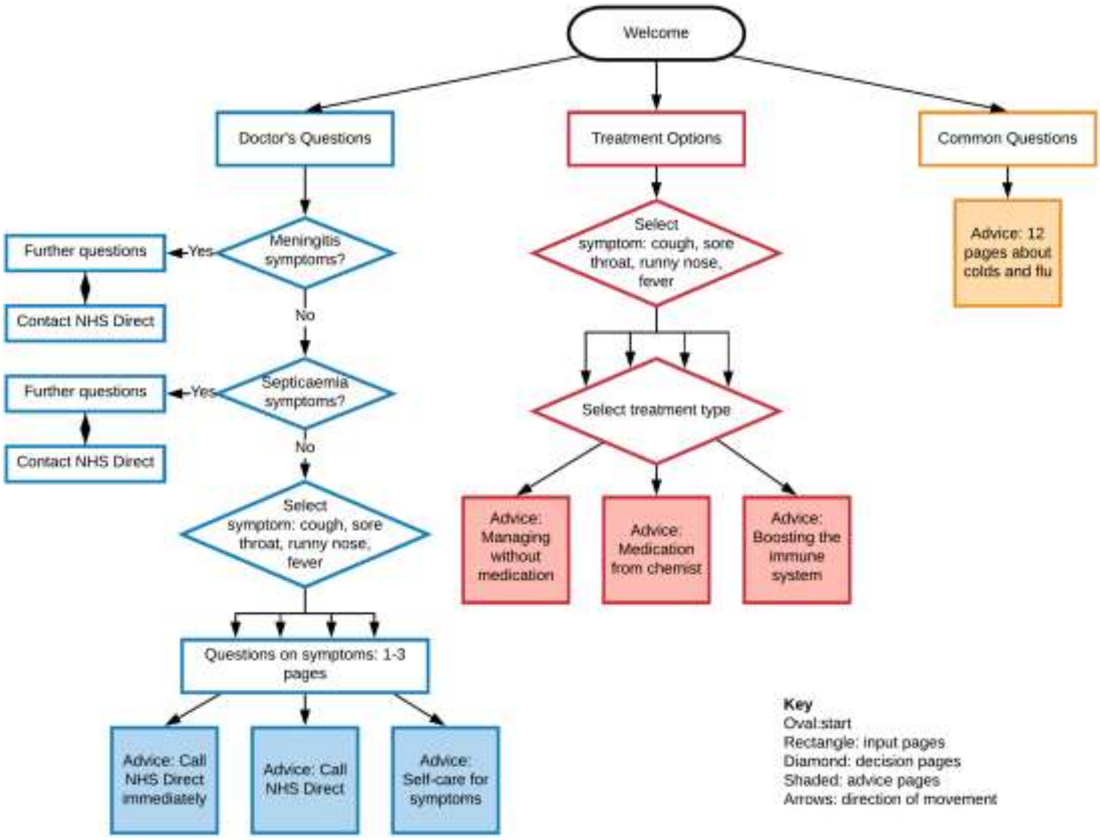


Figure 4. Page flow of the three components of Internet Dr, leading to advice pages

4.3.5.3 Intervention logic model

The logic model illustrates the barriers to self-caring for RTI symptoms that are suggested to influence unnecessary GP visits (see Figure 5). Usage of the theoretically underpinned content was anticipated to change modifiable characteristics underlying those barriers to self-care, leading to increased levels of enablement. For example, the constructs of the Theory of Planned Behaviour (Ajzen, 1991) were measured in order to capture attitudinal and normative beliefs about using and following the intervention advice (see Table 2).

Based on the logic models it is hypothesized that:

H1: Meaningful usage of the intervention is associated with higher levels of enablement.

H2: Baseline user characteristics predict meaningful intervention usage.

H3: Changes in modifiable user characteristics mediate the relationship between meaningful usage and enablement.

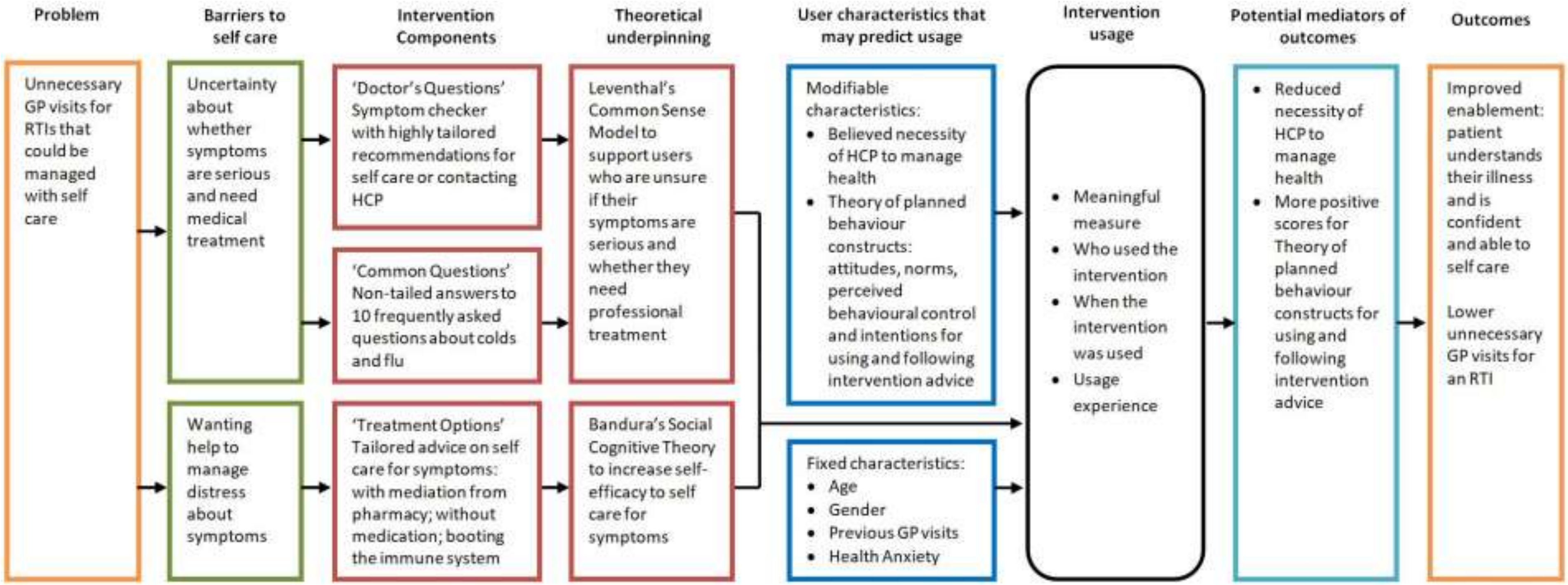


Figure 5. Logic model for Internet Dr intervention.

4.3.6 Statistical Analysis

All analyses examined data from the intervention group only. The LifeGuide Visualisation Tool (Arden-Close et al., 2015) was used to examine which pages had been accessed, at what point in the trial, and the number of users who had viewed them. Statistical analyses were conducted using SPSS for Windows version 24 (IBM Corp). All validated scales were used and scored according to published guidance. Higher scores are indicative of positive change for all measures. Changes in scores for modifiable characteristics were calculated by subtracting individuals' scores at baseline from their scores at follow-up. The frequency distribution of scores for constructs were visually assessed for normality, where these were inconclusive Kolmogorov-Smirnov tests were undertaken. All scales were non-normally distributed, except for change scores. As each advice page was designed to be effective in isolation, as opposed to having an additive effect, a dichotomous categorical variable for users who viewed/did not view advice pages was calculated. All analyses are two-tailed. Due to a data collection error, it was not possible to analyse responses for the TPB construct of attitudes to using and following the advice at baseline and follow-up.

The distribution of scores and residuals for the patient enablement index were positively skewed. Therefore, regression analyses were not possible for PEI. 95% confidence intervals were subsequently carried out to examine mean differences in enablement based on usage, and point biserial Spearman's Rho correlation coefficient to examine the relationships between changes in user characteristics and enablement. Differences in scores for PETS based on usage were also examined using 95% confidence intervals. Logistic regression was used to examine whether user characteristics at baseline predict usage. Simple linear regression analyses were carried out to examine whether usage predicted changes in modifiable characteristics.

4.4 Results

4.4.1 Intervention group characteristics

Thirty-one general practices invited 43,769 patients to take part in the RCT. Of these, 3,044 (6.95%) consented to take part. Over the course of the study 121 participants left their practice, leaving a total of 2,923 users: 1,491 (51.01%) in the intervention group; 1,432 (48.99%) in the control group. Patients' GP notes showed that 269 (18.04%) people in the intervention group had visited their GP for an RTI in the year prior to the study. Over the course of the trial, 852 (57.14%) participants in the intervention group reported having an RTI.

Table 2. User characteristics and psychological measures collected online, and usage data for Internet Dr study.

Measure	Baseline				Follow-up			
	N	M	SD	Min-Max	N	M	SD	Min-Max
Intervention group	1491							
User characteristics								
Age (years)	1490	56.78	13.52	18-89				
Female	816/1490 (54.77%)							
Psychological measures								
Health anxiety	1491	8.20	4.65	0-34				
TPB:								
Subjective Norm	1387	9.07	2.85	0-14	833	8.50	2.72	0-14
PBC	1426	10.20	2.85	0-14	831	8.07	3.63	0-14
Intentions	1445	9.37	3.40	0-14				
Health locus of control	1487	13.90	5.12	0-21	962	13.54	5.05	0-21
Krantz health opinion survey	1490	27.45	9.15	0-49	966	27.35	8.98	0-49
PETS:								
Made symptoms worse					458	4.34	.88	1-5
Uncertain how to use intervention					458	4.37	.95	1-5
Doubts about intervention efficacy					458	4.03	1.08	1-5
Practical problems					458	4.19	.96	1-5
PEI					952	1.86	3.03	0-12

Measure	Baseline				Follow-up			
	N	M	SD	Min-Max	N	M	SD	Min-Max
Summative usage data								
Number of logins					1491	4.86	2.87	0-18
Time spent (minutes)					1491	4.68	6.57	0-44.6
Number of pages viewed					1491	10.10	10.99	0-81
Meaningful usage data:								
Viewed any advice					616 (41.32%)			
Viewed Doctor's Questions advice					244 (16.37%)			
Viewed Treatment Options advice					297 (19.92%)			
Viewed Common Questions advice					372 (24.95%)			

M = mean, *SD* = standard deviation

4.4.2 Describing and defining Usage

4.4.2.1 What type of usage is meaningful to the intervention?

Summative measures of usage for the number of logins, time spent on the intervention and number of pages viewed were examined (Table 2). The number of logins includes completing interim questionnaires every four weeks (when intervention content may not have been viewed). The number of pages viewed and time spent on the intervention varies depending on the size and required interaction for each component (see Figure 4). In addition, the only intervention pages that contained theory-based behavioural change techniques were the advice pages in each of the three components, meaning that only users who viewed an advice page received support to self-care for their symptoms. On this basis, having viewed an advice page from any of the components was considered to be a meaningful way of examining usage.

4.4.2.2 How many people reached advice pages and when were they viewed?

616 (42.32%) of users viewed at least one of the advice pages (see Figure 6 and Table 2). Views by component total 913 indicating that almost half of the 616 users viewed more than one component (n=297, 48%). The level of attrition prior to accessing the three components (n=428, 28.71%) was similar to the attrition within each component prior to reaching an advice page (n=444, 29.78%). Doctor's Questions, the component with the most pages, was accessed by the most users and saw the highest attrition. Of the 616 users who viewed advice pages, the highest proportion (*n*=478, 77.60%) did so during their first login, having just completed the baseline questionnaire.

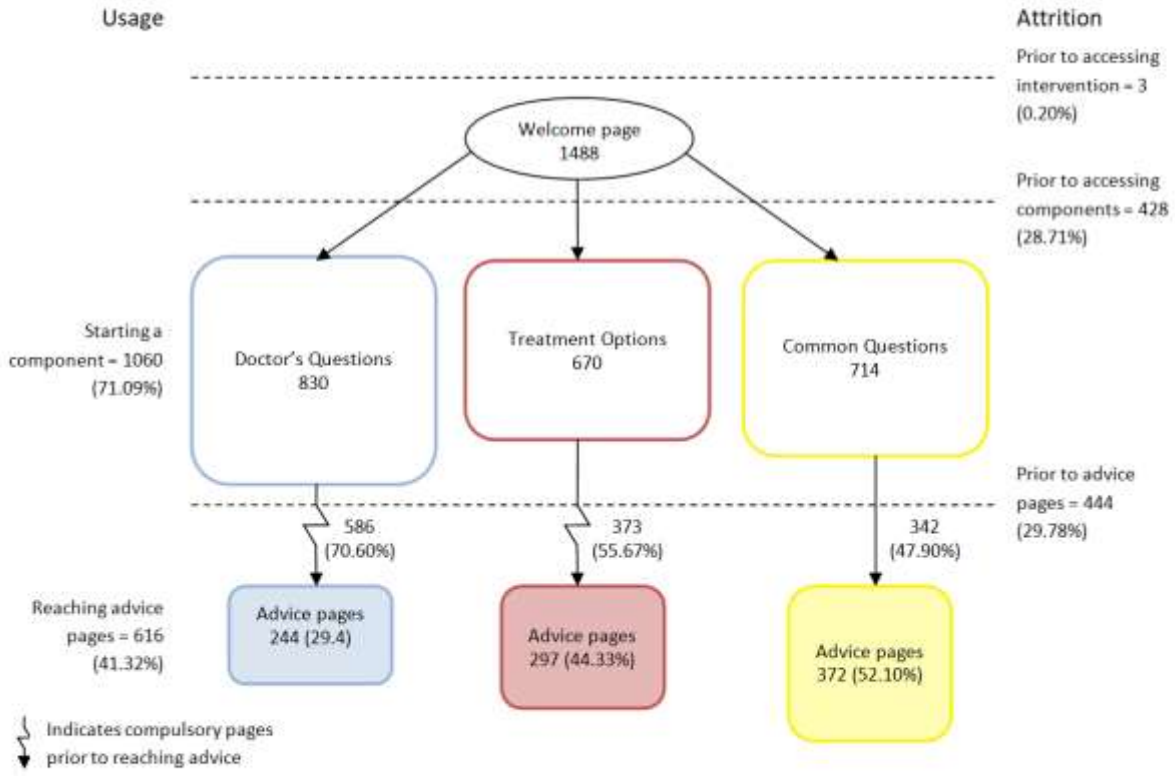


Figure 6. Numbers of users reaching/leaving components and advice pages in Internet Dr intervention

4.4.2.3 Did usage experiences differ for users who viewed/did not view advice pages?

Scores for PETS[24] at 24 weeks revealed that users who had not viewed advice pages were more likely to report that: the intervention made their symptoms worse (*MD* = .26, 95% CI [.10, .42]); they were uncertain about how to use the intervention (*MD* = .29, 95% CI [.12, .46]); they experienced more practical problems that prevented them from accessing the intervention, such as forgetting or being too tired/busy (*MD* = 2.57, 95% CI [.08, .43]). Users who had viewed advice pages and those who had not both held equivalent positive beliefs about the efficacy of the intervention (*MD* = .16, 95% CI [.-.04, .36]).

4.4.3 Testing the logic model

4.4.3.1.1 H1: Viewing advice pages predicts increased enablement

Users who viewed an advice page were more likely to report increased understanding and ability to cope with their illness as a result of having participated in the study, compared to users who did not view advice pages (see Table 3). When examined by individual component, the difference in enablement scores between users who viewed an advice page from Treatment Options and those who did not, was great enough to be practically significant.

Table 3. Confidence intervals comparing scores for enablement at follow-up between users who did/did not view advice pages for Internet Dr intervention.

Component viewed	Not viewed advice pages		Viewed advice pages		Mean difference	95% CI	
	N	Mean	N	Mean		Lower	Upper
Any	532	1.65	420	2.12	.47	.08	.86
Doctor's Questions	785	1.78	167	2.22	.43	-.07	.94
Treatment Options	755	1.68	197	2.55	.88	.40	1.35
Common Questions	683	1.79	269	2.04	.26	-.17	.69

CI = confidence interval

4.4.3.2 H2: Baseline user characteristics predict viewing advice pages.

Users were more likely to view advice pages if they had stronger intentions to use and follow the website advice, and stronger beliefs in their ability to self-manage their illness (see Table 4). The amount of times a user had visited their GP for an RTI in the year prior to the trial was the strongest predictor for viewing advice pages. Other characteristics measured at baseline were not predictive of viewing advice pages.

Table 4. Variables predicting viewing/not viewing advice pages in Internet Dr intervention.

							95% CI	
Predictors	*β*	*SE*	Wald's x^2	*df*	*P*	OR	Lower	Upper

							95% CI	
Predictors	*β*	*SE*	Wald's x^2	*df*	*P*	OR	Lower	Upper
Age	>-.01	>.01	.51	1	.47	1.00	.99	1.01
Gender	-.04	.12	.10	1	.75	.96	.76	1.22
Health anxiety	.02	.01	2.71	1	.10	1.02	1.00	1.05
Theory of planned behaviour:								
Sub. norms	-.01	.02	.13	1	.72	.99	.95	1.04
PBC	.02	.03	.59	1	.44	1.02	.97	1.08
Intentions	.06	.02	7.50	1	.01	1.07	1.02	1.12
Health locus of control	.01	.01	.79	1	.37	1.01	.99	1.04
Krantz health opinion survey	.02	.01	6.75	1	.01	1.02	1.00	1.03
Past GP visits	.30	.08	14.92	1	<.001	1.35	1.16	1.57

SE = standard error, *df* = degrees of freedom, *p* = significance, OR = odds ratio, CI = confidence interval

4.4.3.3 H3: Changes in modifiable user characteristics mediate the relationship between viewing advice pages and enablement.

Viewing advice pages did not predict change in any of the measured modifiable user characteristics over the trial period: norms *F*(1,784) = .12, *P* = .73; PBC *F*(1,798) = 1.09, *P* = .30; health locus of control *F*(1,957) = .14, *P* = .71; Krantz health opinion survey *F*(1,964) = 1.04, *P* = .31. Changes in these modifiable characteristics do not then mediate the observed association between viewing advice pages and enablement.

Strengthened normative beliefs (ie, that friends and family support using and following intervention advice) and increased perceived ease of using and following the advice over the trial period were positively correlated with enablement (r_s = .14, *P* <.001; r_s = .27, *P* <.001 respectively). No relationships were observed between enablement and changes in reported dependence on HCPs (Health Locus of Control r_s = .02, *P* = .47, Krantz Health Opinion Survey r_s = .01, *P* = .91).

4.5 Discussion

This paper presents a process evaluation of data previously collected in a randomised controlled trial of the Internet Dr intervention designed to enable users to appropriately self-care for RTIs

(Little et al., 2016). The aims of the evaluation were to identify meaningful measures of usage (ie, types/patterns of interaction that are relevant to the structure and theory-based content of the intervention) with which to undertake a systematic process analysis, and to examine which parts of the intervention worked, for whom and in what context.

This evaluation provides a clear example of when summative measures of usage (e.g. number of logins, time spent on the intervention) would not provide the fine-grained detail necessary to understand how the intervention worked, and instead identifies usage metrics that are relevant to the structure and theory-based content of Internet Dr. With regard to the number of logins, these include users who logged in to complete interim study questionnaires as well as users who were accessing the intervention. Therefore, inferential analyses using the number of logins would capture users' engagement with the trial as well as with the intervention. Alternatively, if the number of pages viewed had been analyzed and found to be associated with increased enablement it is unclear what this would have meant or how it may be applied to improve the intervention. Each of the components varied in size, so that users who chose the Doctor's Questions component may have viewed five pages and then left the component without having reached any advice on self-caring (see Figure 4). In comparison, if users who visited the Common Questions component also viewed five pages, they would have been able to access four pages of advice. Without knowing what the content of the pages are, the number of pages viewed provides little insight in to how users experienced increases in enablement or for improvements to the website. Instead, as the advice pages are the only content aimed at supporting users to self-care for their illness, viewing advice pages from the different components were identified as the most meaningful metrics with which to analyse usage of the intervention. Users who viewed any advice page were more likely to report higher levels of enablement at 24 weeks compared to users who did not. Although this effect was fairly small, when analyzed by individual component it was apparent that viewing an advice page from Treatment Options led to a practically significant increase in enablement. Therefore, viewing an advice page represents the minimal type and amount of usage required to improve outcomes, and may be described as 'effective engagement' with the intervention (Yardley et al., 2016).

The three components were initially accessed by similar volumes of users, with almost half accessing more than one component. This suggests that offering a variety of content may be useful to maximize the number of users who reach key features of an intervention. The majority of users (n=478) who viewed advice pages did so during their first login, having just completed the baseline questionnaire. Given that this figure represents more than half of all users in the intervention group who reported having an RTI during the 24 week trial (n=852), it is unlikely that everyone accessing advice was experiencing an RTI at that precise time. Most, if not all, users will have experienced RTIs in the past and be familiar with any symptoms they find challenging. This prior experience could have supported 'well' users to engage fully with the intervention and access pertinent advice without having to experience the symptoms at that point in time. This would also be the case for many other common ailments, which suggests that promoting intervention usage when well would be effective for self-caring for future minor ailments, as seen with RTIs.

As past behaviour is typically a strong predictor of future behaviour (Ouellette & Wood, 1998) and previous GP visits reinforce a patient's decision to return to the GP in future (Banks, 2010), it was anticipated that users who had consulted their GP for an RTI in the year prior to the study would be less likely to use the intervention and view advice pages. However, the process analysis shows the reverse to be true, with these target users being more likely to view advice pages. As using the Internet Dr has already been shown to lower the number of GP visits (Little et al., 2016), it is probable that reaching these users was key in achieving that.

Viewing advice pages is important for increasing users' enablement to self-care, with Treatment Options advice showing greatest impact. However, users' characteristics measured across the study provide only a limited explanation of the psychological changes that led to better enablement. Internet Dr content is underpinned by Social Cognitive Theory (Treatment Options) and Leventhal's Common Sense Model (Doctor's Questions, Common Questions). The TPB was selected, along with measures of beliefs in necessity of HCPs to manage illness, to measure psychological changes across the study. The construct of perceived behavioural control (PBC) within the TPB, was designed by Azjen (2002) to capture self-efficacy using the definition laid out previously by Bandura. Therefore, users who accessed Treatment Options content based upon Social Cognitive Theory were expected to report increased PBC, yet this effect was not found. Bandura (2006) stipulates the use of measures of self-efficacy with social cognitive theory, and considering that PBC encapsulates several constructs (Trafimow, Sheeran, Conner, & Finlay, 2002), a general self-efficacy measure may be preferable for future studies of this nature (Scherbaum, Cohen-Charash, & Kern, 2006). Since the development of Internet Dr in 2008, further research has identified additional barriers to self-care (e.g. perceptions of illness severity, not considering alternative options, cost implications of paying for un-prescribed medication; Richardson et al., 2018). Although these were not intentionally targeted within Internet Dr, the increases in enablement and the lower GP visits suggest the theoretically-underpinned content may have been effective at addressing some of these additional barriers. Future research including measures of user characteristics that better reflect these barriers may provide a more in-depth explanation of the association between usage and enablement.

As well as examining users who viewed advice pages, it is also important to consider that almost 60% of users in the intervention group did not access any advice pages. The first point of attrition for these users occurred at the ‘Welcome Page’, with just under a third of users not progressing any further (see Figure 6). One explanation for this may be that many of these users were not ill at this time and intended to return to the intervention should they experience an RTI. However, these findings suggest that using the advice pages when well can still be of benefit. Users who did not view advice pages were also more likely to report practical barriers to usage (e.g. too tired or busy, forgetting). Applying these findings to future dissemination means that patients can be encouraged to access the intervention at a time that is convenient to them, and to not have to wait until they are experiencing symptoms. This could be reiterated to users by adding a message to the Welcome Page.

The second incidence of attrition occurred within the components, with almost another third of users starting a component but not reaching an advice page. The highest proportion of this attrition occurred in the Doctor’s Questions component, which is the largest component. This section includes a compulsory symptom checker with up to 10 pages of questions, necessary in any intervention for minor ailments to ensure that serious infections in need of urgent medical attention are identified. Although these questions could not be omitted, the format and layout could be amended so as to minimize burden for the user and subsequent attrition. For example, streamlining content by combining pages or motivating users to continue by including page numbers or breadcrumbs to show progress and location may have lowered attrition. Interestingly, these compulsory pages were not raised as a concern in qualitative evaluation carried out during the development phase (Yardley et al., 2010). This highlights the ability of process evaluations to establish design precedents from post-hoc data analysis such as ensuring that users are motivated and able to access the active ingredients of interventions with minimal effort.

4.5.1 Limitations

Scores for increased enablement were low with the majority of users selecting ‘same or less’ (0), which probably reflects that most users have experienced and successfully self-cared for RTIs previously. This is supported by the finding that users who had failed to self-care prior to the study were more likely to use the advice pages. The PEI scoring was problematic as there were only three response options available which did not allow users to distinguish between no change and deterioration in enablement. The resultant skew in scores meant that regression analyses were inappropriate for examining variables predicting enablement. Recent studies have measured the PEI using likert scales of five or more, allowing for multiple, finer-graded levels of response (Everitt et al., 1613; Morrison et al., 2018).

The purpose of this study was to examine the psychological outcomes for intervention users, and to explain how the intervention had been effective. As a result of the problematic PEI scoring, we considered examining the relationship between usage of advice pages and the behavioural outcome of GP visits. However, this was not possible for several reasons. Almost half the users who had viewed an advice page accessed advice from more than one component, and some users viewed several advice pages within a component, meaning that it was not possible to match the advice received to the action taken. In addition, most users had accessed advice prior to becoming ill. Finally, users who were advised to contact NHS Direct may have been recommended by NHS Direct to contact their GP, but this was not captured in either self-report data or patients' GP notes.

Initial RCT uptake by patients was just under 7%, suggesting that participants were more willing to engage in this type of research and may not be representative of the wider population. Therefore, participant and non-participant characteristics were compared for the RCT analysis, and as the index for multiple deprivation showed that participants were less deprived than the wider population, the RCT results were controlled for this variable. However, as the process evaluation is a secondary analysis of the data collected in the RCT, identifiable details (i.e. home address) were removed from the usage data, as a consequence it was not possible to control for possible effects that lower levels of deprivation may have had on behavioural engagement.

The Internet Dr was developed and trialled over 10 years ago, and during this time digital intervention technology has advanced considerably. In addition to the recommendations from the usage analysis, before further dissemination activities are undertaken the intervention would need further testing and developing to ensure that current accessibility guidelines are met and the content is mobile-friendly.

4.5.2 Conclusions

The findings from the process evaluation demonstrate the advantages of using systematic methods for analysing digital intervention usage. By identifying specific metrics that are meaningful to the intervention structure, theory-based content and target behaviour it was possible to examine how the intervention was effective, for whom and in what context, and to provide specific recommendations for improving intervention design and implementation. Inferential analyses of usage identified that viewing advice pages from the Internet Dr Intervention is effective at increasing enablement to self-care for the symptoms of RTIs. Having identified content that is crucial for behaviour change, this provides the opportunity to ensure that prior compulsory pages are streamlined to maximize the number of users reaching these 'active ingredients', thereby minimizing attrition. However, streamlining within components does not necessarily mean reducing the amount of components available as users utilized the choice. These findings suggest that viewing advice pages prior to having an RTI encouraged users to self-care for future symptoms. This means that for Internet Dr's dissemination, users may be encouraged to access the intervention at their convenience rather than wait for the occurrence of an illness. The intervention was effective at reaching the target population of users who had previously failed to self-care for their symptoms and consulted their GP. Taking these findings in to consideration, the Internet Dr provides a model for future digital interventions aiming to increase self-care for other minor ailments.

4.5.3 Acknowledgements

We would like to thank the Economic and Social Research Council for funding this paper (award: 1692136 – Understanding digital intervention engagement: Making sense of large-scale usage data). The Internet Dr development and randomised controlled trials were funded by the National Institute for Health Research Programme Grants for Applied Research (grant ref No RP-PG-0407-10098). The Internet Dr website was developed using LifeGuide, which was partly funded by the NIHR Biomedical Research Centre (BRC), Southampton. LY is a National Institute for Health Research (NIHR) Senior Investigator of the NHS, the NIHR, or the Department of Health. LY is affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Behavioural Science and Evaluation of Interventions at the University of Bristol in partnership with Public Health England (PHE).

4.5.4 Conflicts of Interest

None declared.

Chapter 5 Paper 3: Process evaluations of a web-based intervention to increase handwashing during a pandemic: Moving Germ Defence from a randomised controlled trial to public dissemination

Miller, S., Ainsworth, B., Weal, M., Smith, P., Little, P., Yardley, L., & Morrison, L. (2021). Process evaluations of a web-based intervention to increase handwashing during a pandemic: Moving Germ Defence from a randomised controlled trial to public dissemination. *Journal of Medical Internet Research,* 26104 (forthcoming/in press).

5.1 Abstract

Background: Washing hands helps prevent transmission of seasonal and pandemic respiratory viruses. The PRIMIT study developed a fully automated, digital intervention to promote handwashing. In a randomised controlled trial during the Swine Flu outbreak, participants who had access to the intervention reported washing their hands more and experienced less respiratory tract infections than those without access. Using these findings, the intervention was subsequently adapted, renamed 'Germ Defence', and a study designed to assess a preliminary dissemination of the intervention to the general public to help prevent the spread of seasonal colds and flu.

Objectives: This paper compares process evaluations of the PRIMIT trial and Germ Defence dissemination to examine: 1) how online research enrolment procedures impacted on who used the intervention; 2) intervention usage in the two contexts; 3) whether increased intentions to wash hands are replicated once disseminated. The purpose of this paper is to provide insight so that the Germ Defence intervention could be optimized for wide-scale dissemination in the event of a global pandemic.

Methods: The PRIMIT trial ran between 2010 and 2012 recruiting participants offline from General Practices, with restricted access to the intervention (N=9155). Germ Defence was disseminated as an open access website for use by the general public from 2016 to 2019 (N=624). The process evaluation plan was developed using Medical Research Council guidance and the framework for Analysing and Measuring Usage and Engagement Data. Both interventions contained a goal-setting section where users self-reported current and intended handwashing behaviour across seven situations.

Results: During online enrolment, 54.30% (n=17,511) of PRIMIT study participants dropped out of the study compared to 36.46% (n=358) of Germ Defence users. Having reached the intervention, 93.79% (n=8586) of PRIMIT users completed the core section, whereas 65.06% (n=406) of Germ Defence users reached the same point. Users across both studies selected to increase their handwashing in five out of seven situations, including before eating snacks (PRIMIT *MD*=1.04 [*CI* 1.02, 1.06], Germ Defence *MD*=.95 [*CI* .77, 1.13]) and after blowing their nose, sneezing or coughing (PRIMIT *MD*=.99 [*CI* .97, 1.02], Germ Defence *MD*=.84 [*CI* .68, 1.01]).

Conclusions: By comparing a preliminary dissemination of Germ Defence to the PRIMIT trial we have been able to examine the potential effects of research procedures on uptake and attrition, such as the sizeable dropout during the PRIMIT trial enrolment procedure that may have led to a more motivated sample. The Germ Defence study highlighted points of attrition within the intervention. Despite sample bias in the trial context, the intervention replicated increases in intentions to handwash when used ‘in the wild’. This preliminary dissemination study informed the adaptation of the intervention for the COVID-19 health emergency, and it has now been disseminated globally.

Trial Registry: ISRCTN75058295

5.2 Introduction

Pandemic respiratory viruses present a global health threat, leading to more deaths across a wider spread of the population than seasonal flu, as seen through outbreaks such as SARS-CoV (Severe Acute Respiratory Syndrome), Swine Flu (H1N1 influenza) and MERS-CoV (Middle East Respiratory Syndrome), as well as COVID-19 (SARS-CoV-2) (World Health Organization, 2010). However, for more vulnerable groups such as the elderly or seriously ill, seasonal flu can still present a serious health risk and increased likelihood of needing medical care (World Health Organization, 2010). Handwashing is an accessible and simple infection control behaviour (Morrison & Yardley, 2009; Yardley, Miller, Teasdale, & Little, 2011) that has been promoted to the general public for many decades to slow the spread of cold and flu viruses both seasonally and during pandemics (Global Handwashing Partnership, 2020; World Health Organization, 2009). The COVID-19 pandemic has seen renewed calls for increased handwashing from governments and health organisations around the world as a means to control the spread of the virus (Centers for Disease Control and Prevention, 2020; Department of Health and Social Care, UK Government, 2020). Yet despite high levels of health promotion and public awareness of handwashing, evidence from prior pandemics (i.e. SARS-CoV and Swine flu) suggests that increases in reported levels of handwashing were low (Lau, Kim, Tsui, & Griffiths, 2007; Rubin, Amlôt, Page, & Wessely, 2009). An effective, evidence-based behavioural intervention to support increased handwashing within the home in the event of pandemics such as COVID-19, but also for seasonal respiratory tract infections (RTIs), is urgently required (Gold et al., 2020; West, Michie, Rubin, & Amlôt, 2020; Wu, Riley, Fraser, & Leung, 2006). Digital interventions provide the advantages of being quick to disseminate and flexible so that content can be updated to reflect changes during a pandemic.

PRIMIT (A PRImary care trial of a website based infection control intervention to Modify Influenza-like illness and respiratory infection Transmission) was commissioned and funded by the UK Medical Research Council (MRC; 2008) for use in the event of a pandemic (Little et al., 2015). A theory-based, stand-alone, digital intervention to increase handwashing targeting the general population was developed using the Person-Based Approach (Yardley et al., 2015). This included extensive mixed-methods evaluation of the usability, functionality and acceptability of the intervention (Miller, Yardley, & Little, 2012; Yardley, Miller, Teasdale & Little, 2011; Yardley, Miller, Schlotz, & Little, 2011). A randomised controlled trial (RCT) of over 20,000 participants was carried out in the UK whilst Swine Flu was circulating in the community. The findings established that: participants who had access to the intervention reported washing their hands more than those who did not; they, and the people they lived with, had less RTIs; users who contracted RTIs were ill for less time (Little et al., 2015). A behavioural analysis of the data collected during the RCT demonstrated that viewing each of the four available sessions led to additive increases in handwashing levels (Ainsworth et al., 2016). However, completing the first session was deemed to be 'effective engagement': after viewing this session the majority of users had increased their handwashing to sufficient levels to lower transmission of viruses (Ainsworth et al., 2016; Yardley et al., 2016). This session contained a range of motivational messages, and included a goal-setting behaviour change technique (BCT). The goal-setting section required users to consider their current handwashing frequency across a range of specific situations (e.g. after going to the toilet, before eating a meal), and then make a plan to increase handwashing in the future. Based on the findings from the behavioural analysis (Ainsworth et al., 2016), in 2016 the architecture of the intervention was adapted to enable access to all the intervention content in one session (rather than four sequential sessions). The intervention was renamed 'Germ Defence' and disseminated to the general public to lower transmission of seasonal cold and flu viruses (Miller et al., 2019). The data generated by the PRIMIT and Germ Defence studies provide us with the opportunity to compare uptake and usage of the intervention in the RCT and 'in the wild' contexts.

RCTs are considered to be the 'gold standard' for evaluating intervention efficacy. However, they have been criticized for producing results that are not replicated once an intervention is freely disseminated to the intended population (Clay, 2010; Deaton & Cartwright, 2018; Fleming et al., 2018). RCTs require artificial conditions that differ from how an intervention may be used 'in the wild'. For example, RCTs of digital interventions typically involve effortful online research procedures such as lengthy enrolment processes, followed by in-depth baseline, interim and follow-up data collection across several weeks, if not months (Michie et al., 2017). These demanding procedures may affect both the type of people who take part in an RCT and how they use the intervention (Fleming et al., 2018; Tripepi, Jager, Dekker, & Zoccali, 2010; Van Der Mispel, Poppe, Crombez, Verloigne, & De Bourdeaudhuij, 2017). Participants with lower levels of education or health literacy are more likely to drop out of a study during these processes (Reinwand et al., 2015), meaning that these groups are unintentionally excluded. Conversely, this can lead to volunteer bias within the sample (Tripepi et al., 2010), whereby people who are more highly motivated to perform the behaviour, or have higher levels of health literacy/education are more likely to participate in the trial. In addition, research suggests that effortful procedures may increase participants' sense of support or accountability, leading to artificially high levels of engagement (Fleming et al., 2018).

Calls have been made to evaluate digital interventions so that their effectiveness may be established beyond trial conditions (Michie et al., 2017). However, a review of studies of publicly available digital interventions found that few reported in-depth usage data, and only one intervention was identified as having been empirically examined through an RCT as well (Fleming et al., 2018). Having established the effectiveness of the PRIMIT intervention in a trial context, the Germ Defence study was subsequently devised as a novel preliminary dissemination of the intervention to the general public. Comparing intervention usage across these two contexts provides the opportunity to examine the impact of the rigors of RCTs on intervention usage and effectiveness, and to assess whether the intervention maintains efficacy when accessed 'in the wild' (Glasgow, Vogt, & Boles, 1999; Murray et al., 2010).

This paper reports process evaluations (Moore et al., 2015) of the PRIMIT and Germ Defence studies. The aim of this study was to provide insights that would help to optimize the intervention and maximize reach in the event of a pandemic, and to provide a comparison between RCT and 'in the wild' contexts. This has proved timely, as the Germ Defence intervention has now been adapted and widely disseminated for use during the COVID-19 pandemic (www.germdefence.org; Ainsworth et al., 2020). By comparing trial and 'in the wild' contexts we were able to examine the effect of the RCT research procedures on uptake and usage. Examining attrition during trial procedures has the potential to inform RCT design and produce findings that will translate when implemented at scale within the community. Also, examining usage of intervention content across the two contexts provides insight in to how to design interventions that are likely to be more effective and engaging when disseminated at scale. As the intervention is effective where other public health campaigns have had less success (Lau et al., 2007; Rubin et al., 2009), it is important to understand how increased handwashing was achieved. The Germ Defence study enabled us to examine whether the goal-setting section continued to support increased intentions for handwashing in specific situations when the intervention is not supported by the trial context.

The specific objectives were to compare:

- Completion of the online research enrolment procedures
- Usage of the intervention content
- Self-reported intentions to wash hands in specific situations

5.3 Methods

5.3.1 Study procedures

The PRIMIT trial ran across three winters from 2010 to 2012. Participants were randomised equally to either intervention or control groups. Both groups were required to give consent, login and complete baseline measures to enrol in the study. Participants in the intervention group were given immediate access to the first session of the website (see Figure 7). The trial is registered with the ISRCTN registry, number ISRCTN75058295, for full details of the trial see Little et al., 2015.

Germ Defence was available from 2016 to 2019. On the web-pages prior to accessing the intervention, users were provided with details of the study and asked if they would like to participate by providing baseline measures or go directly to the intervention. For users volunteering to complete baseline measures, the consent process was adapted to provide minimal but sufficient information for informed consent, with full details available as an optional click-through.

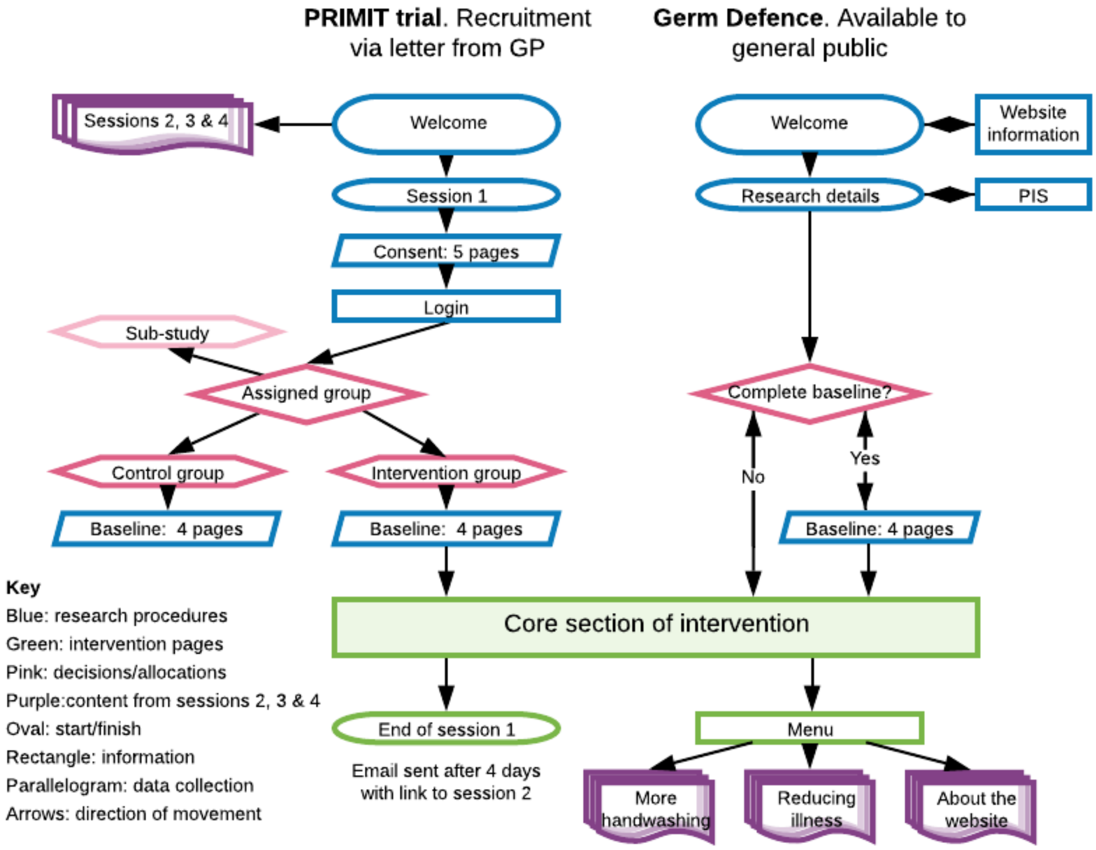


Figure 7. Research procedures for PRIMIT and Germ Defence studies

5.3.2 Design

A process evaluation plan to compare usage of PRIMIT and Germ Defence studies and interventions was developed whilst adapting and preparing Germ Defence for dissemination in 2016. The plan was structured using the AMUsED Framework (Miller et al., 2019) (see Appendices B-D) and is in line with MRC process evaluation guidance (Moore et al., 2015). The process evaluation team combined expertise in psychology, primary care, statistical analyses and computer science. Most team members were already familiar with the intervention having worked on the PRIMIT study development (Miller et al., 2012; Yardley, Miller, Schlotz, & Little, 2011; Yardley, Miller, Teasdale, & Little, 2011) or RCT (Ainsworth et al., 2016; Little et al., 2015).

5.3.3 Participants

PRIMIT study participants were recruited to the RCT by letters sent from National Health Service (NHS) general practices (GPs) to patients aged 18 years or older, living with at least one other person who was willing to report their illnesses for the study, and had internet access. Patients with severe mental problems, who were terminally ill or who had a skin complaint that restricts handwashing were excluded from the study. Access to the website was restricted so that only users who received specific login details were able to enrol and consent to the study online. Participants were not compensated for taking part in the study.

The Germ Defence study was accessible to members of the general public who indicated being aged sixteen or over. Website details were distributed to various health sector organisations for use with their clients (e.g. health support charities for people at risk from RTIs and local council public health organisations). Germ Defence is an endorsed resource for the National Institute for Health and Care Excellence (NICE) antimicrobial resistance (AMR) guidelines and a link to the intervention is provided on the NICE AMR Guidelines website. The intervention was also promoted directly to the general public (e.g. through a chain of high-street pharmacies, play groups, social media and student intranet).

5.3.4 Ethics

The PRIMIT study was approved by a multicentre research ethics committee (number 08/H0502/14). The Germ Defence study was approved by the University of Southampton, School of Psychology ethics committee (number 19399).

5.3.5 Measures

The PRIMIT study and Germ Defence websites were developed using LifeGuide software (Williams et al., 2010). The software collected self-report measures and automatically recorded user interaction with the intervention such as time, date, pages viewed and the order they were viewed in.

For the Germ Defence study, web-based, self-report baseline measures of user characteristics and behaviours were streamlined in comparison to those used in the PRIMIT study (see Table 5). Measures were selected based on the findings from the PRIMIT study (Little et al., 2015) and completion of the AMUsED framework checklists (Miller et al., 2019). By making the baseline measures voluntary, it was possible to split the sample in to two groups: those who chose to complete these measures, and those who did not.

Table 5. Germ Defence and equivalent PRIMIT online study measures.

Measure	**Description**	**Response options**	**PS**	**GD**
Baseline online self-report questionnaires				

Measure	Description	Response options	PS	GD
Gender		Male, Female, Other	✓	✓
Age	Select age from dropdown box.	PS: 18-99, GD: 16-99	✓	✓
Recruitment	How did you hear about the website?	health support group, school or children's group, search engine, social media, news article, word of mouth, other.		✓
Current daily handwashing	How many times a day do you wash your hands, including using antibacterial hand gel, on average?	0-2, 3-4, 5-6, 7-9, 10 or more.	✓	✓
Intended daily handwashing	How many times a day do you intend to wash your hands in the future, including using antibacterial hand gel, on average?	0-2, 3-4, 5-6, 7-9, 10 or more.	✓	✓
Perceived likelihood of users catching an RTI	PS: 2 items (e.g. My chances of catching a cold or flu are high if I don't take action to prevent it).	1-7 Likert scale ranging from 'strongly disagree' to 'strongly agree'	✓	✓
	GD: My chances of catching a cold, flu or stomach bug are high if I don't take action to prevent it.	1-7 Likert scale ranging from 'strongly disagree' to 'strongly agree'.		
Perceived severity of catching an RTI	If I catch a cold, flu or stomach bug I am likely to become seriously ill.	1-7 Likert scale ranging from 'strongly disagree' to 'strongly agree'.		✓
Goal setting section within intervention				
Current behaviour	Over the last week I washed my hands:			

Measure	Description	Response options	PS	GD
	C1: Before I ate a meal	1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always'	✓	✓
	C2: Before I ate snacks	1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always'	✓	✓
	C3: After I went to the toilet	1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always'	✓	✓
	C4: When I came in to the house	1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always'	✓	✓
	C5: After I had been close to someone with a cold, flu or upset stomach	1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always' GS only: 0='not applicable'.	✓	✓
	C6: After blowing my nose, or sneezing/coughing on my hands	1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always' GD only: 0='not applicable'	✓	✓
	C7: After touching something with lots of germs on	GD only: 1='almost never', 2='sometimes', 3='quite often', 4='very often', 5='almost always' 0='not applicable'		✓
Intended behaviour	I will try to wash my hands: Items as for current behaviour	Responses as for current behaviour		

PS is PRIMIT study, GD is Germ Defence

5.3.6 Intervention structure

The intervention used for the PRIMIT and Germ Defence studies was developed by researchers at the University of Southampton. Full details of intervention structure and development are reported elsewhere (Little et al., 2015; Miller et al., 2012; Yardley, Miller, Schlotz, & Little, 2011; Yardley, Miller, Teasdale, & Little, 2011) and archived copies available at http://archive.germdefence.org/ (see PRIMIT and Germ Defence v1). Intervention content for both studies was “frozen” during the trial and dissemination periods.

PRIMIT study users had access to four sessions of theory- and evidence-based content released sequentially over three and a half weeks. Having established that viewing the first session of the PRIMIT study intervention represented effective engagement (Ainsworth et al., 2016), Germ Defence was also structured with the same content as the core section (see Figure 8). After completing the core section, all information from the following three sessions were then available, allowing users to access the whole intervention in one visit, with no further sessions necessary. These core pages focused on informing, supporting and motivating users to wash their hands more (Miller et al., 2012; Yardley, Miller, Schlotz, & Little, 2011) and were ‘tunnelled’ so that users had to move through pages in a specific order. The appearance was updated for Germ Defence, but only two content changes were made to the core section: details about pandemic flu (Miller et al., 2012) and particularly the Swine Flu outbreak that started in 2009 were removed; a motivational message from the original ‘session 2’ was added in between the goal setting pages.

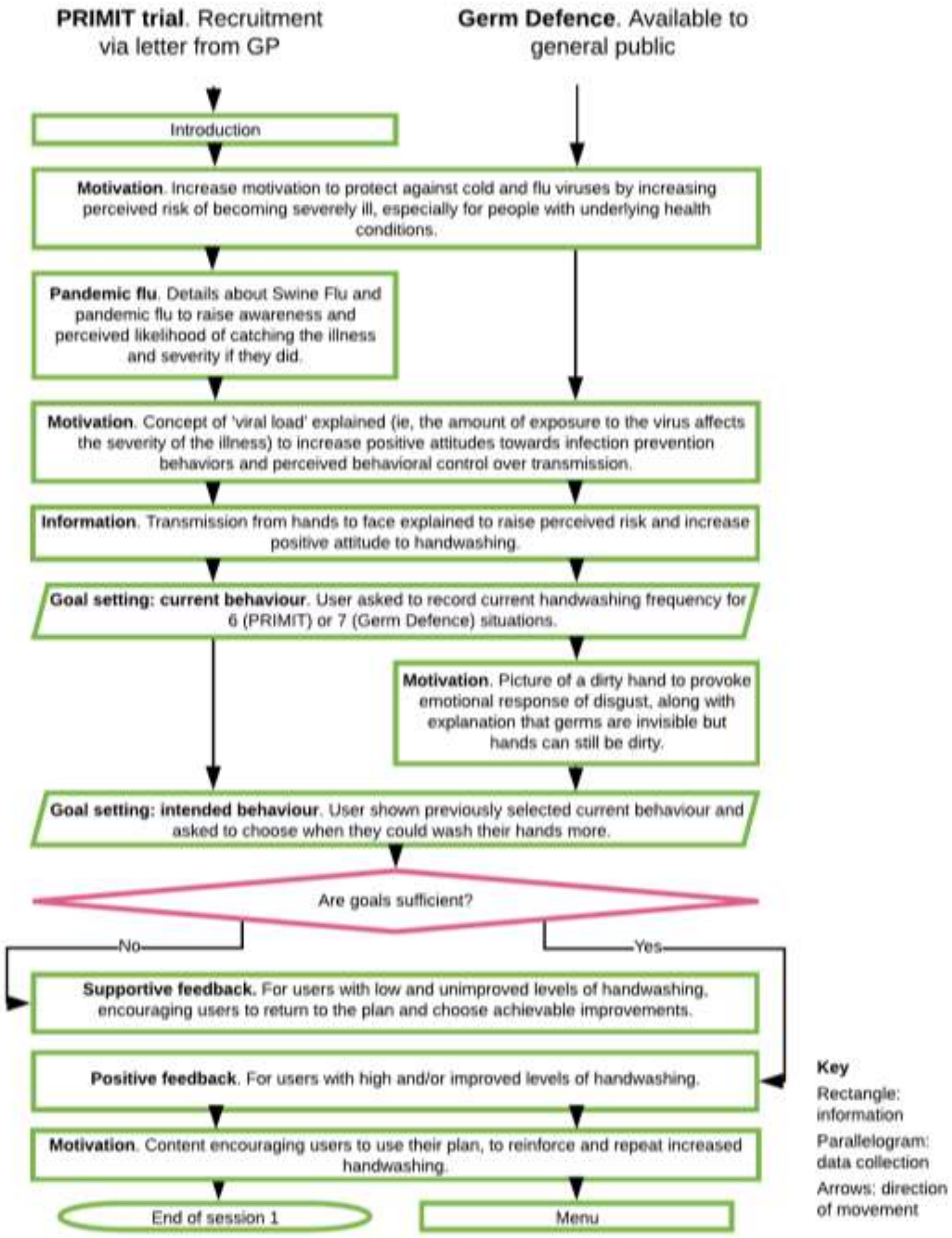


Figure 8. PRIMIT study and Germ Defence intervention structure for core section.

Within this core section was a goal setting BCT encouraging handwashing in situations where there is increased risk of infection through spreading the virus inside the home (e.g. when I come in to the house) or moving viruses from hands to face (e.g. before I eat snacks). These are the only pages in the intervention that require user interaction, and users were only able to progress further if they selected current and intended future handwashing for all situations. Users were provided with feedback after completing their plan: positive for high or improved levels of handwashing; supportive for low and unimproved levels of handwashing suggesting users revisit and change their plan. Having finished these pages, PRIMIT study users had completed session 1.

For Germ Defence users, after this core section was a menu allowing users to access three further components with information from sessions 2-4 of the PRIMIT study. These included further support for handwashing and also information on other infection prevention behaviours (i.e. social distancing, not touching your face, wearing a mask, cleaning surfaces).

5.3.7 Data analysis

Anonymous usage data collected automatically through interaction with the intervention was examined for all Germ Defence users, and PRIMIT participants in the intervention group. Usage of the core section of Germ Defence is reported for the whole sample and the sub-sample of users who completed baseline measures. Responses to self-report measures (i.e. at baseline and goal-setting) were analyzed for the PRIMIT study intervention group and Germ Defence baseline respondents. As part of the PRIMIT trial requirement, a small proportion of users were placed in to a sub-study and completed different baseline measures that excluded handwashing behaviours. As this group cannot be included in analyses with other users, they have been removed from this process evaluation.

SPSS for Windows version 25 (IBM Corp) was used for all statistical analyses. The frequency distribution of scores for self-report measures was visually assessed for normality and all measures of user characteristics at baseline were non-normally distributed. All analyses were two-tailed. Due to the difference in sample sizes, analyses of usage data are reported as percentages to enable more meaningful comparison.

5.3.7.1 Completion of online research enrolment procedures

To examine completion of the online research enrolment procedures, usage data were analyzed for proportions of users who had: viewed each page; continued on to another page; left the study. Due to changes in data collection across the three winters, usage data covering research enrolment for the first two winters are not available (intervention usage data were unaffected). Therefore reported usage percentages for enrolment pages are taken from users recruited in the third winter, representing over 75% of the total sample. The number of views for the welcome pages (PRIMIT study n=209,852, Germ Defence n=12,106) include multiple views by the same participants, people who arrived at the page unintentionally, and bots (automated software programs). In addition, PRIMIT study participants who were returning to the intervention to view later sessions are included in this number. To enable equivalent comparison between the two research procedures and ensure that usage was intentional, the first page has been separated and proportions of participants have been calculated based on the sample who reached the second page.

5.3.7.2 Intervention usage

Intervention content usage was examined for pages viewed and attrition in the core section of the intervention. Confidence intervals at 95% were used to compare means of baseline behavioural and psychological measures to determine whether there were any differences between intervention users in the two contexts. Logistic regression analyses were carried out to examine whether baseline user characteristics predicted completion of the core section of the intervention.

5.3.7.3 Handwashing behaviours

Confidence intervals at 95% were also used to compare scores for intentions to handwash in the goal-setting situations. The first completion of intended handwashing responses for the goal-setting items was analyzed. Confidence intervals comparing individual handwashing situations are reported for practical difference, with mean differences >0.3 on a scale of 1-5. Paired sample cases items were excluded listwise to reduce any bias due to attrition, and equal variances were assumed for independent samples.

5.4 Results

5.4.1 Completion of online research enrolment procedures

Letters inviting patients to take part in the PRIMIT study were sent to 804,897 individuals, of which 2.49% (n=20,042) provided online consent, completed baseline measures and were assigned to a group (see Figure 9). The Germ Defence website received 12,106 visits to the first page. The majority of these were identified as bots 63.1% (n=7639). Of the remaining 4467 visits: 4.99% (n=223) continued to provide baseline measures and access the intervention; 8.98% (n=401) accessed the intervention directly.

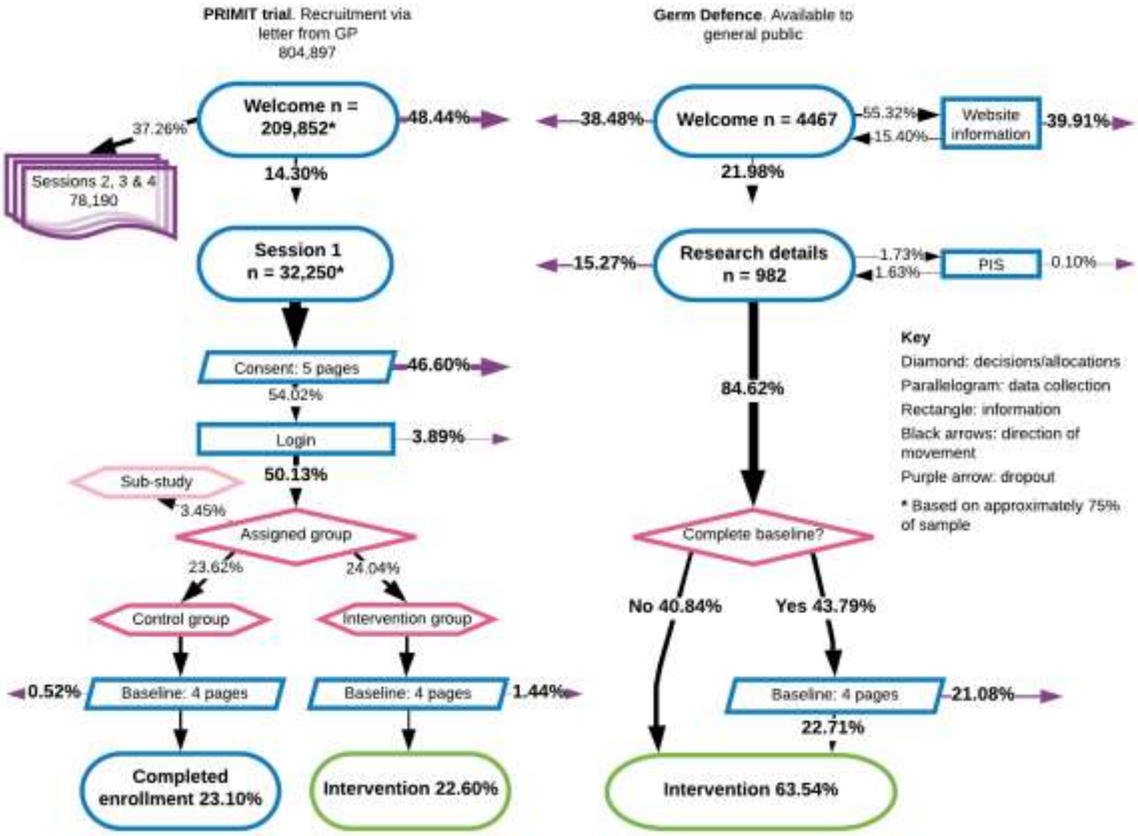


Figure 9. Flowcharts of usage during PRIMIT study and Germ Defence research procedures.

5.4.2 Intervention usage

5.4.2.1 Who used the intervention?

The Germ Defence sample was younger (M=42 years) and predominantly female (75.72%) compared to PRIMIT (M=57 years, 55.93%). The samples were similar in most of the self-reported behaviours and beliefs at baseline (see Table 6). PRIMIT study users reported higher intended levels of daily handwashing at baseline (*M*=3.97) compared to current daily levels (*M*=3.85)(*MD*=.12, 95% CI [.11, .13]), as did Germ Defence users (intention *M*=4.11, current *M*=3.87) (*MD*=.24, 95% CI [.13, .35]). Comparisons of scores between PRIMIT study and Germ Defence users for current (*MD*=-.022, 95% CI [-.17, .13]) and intended (*MD*=-.15, 95% CI [-.30, -.01]) daily handwashing showed no practical differences. Germ Defence users perceived themselves as more likely to contract an RTI (*M*=5.53) than the users in the PRIMIT study intervention group (*M*=5.11)(*MD* = .42, 95% CI [.20, .64]). None of the user characteristics recorded at baseline (e.g. age, daily handwashing behaviour) predicted completion of the core section for either PRIMIT or Germ Defence (see Appendix I).

Table 6. Characteristics of PRIMIT study intervention group and Germ Defence study baseline responders

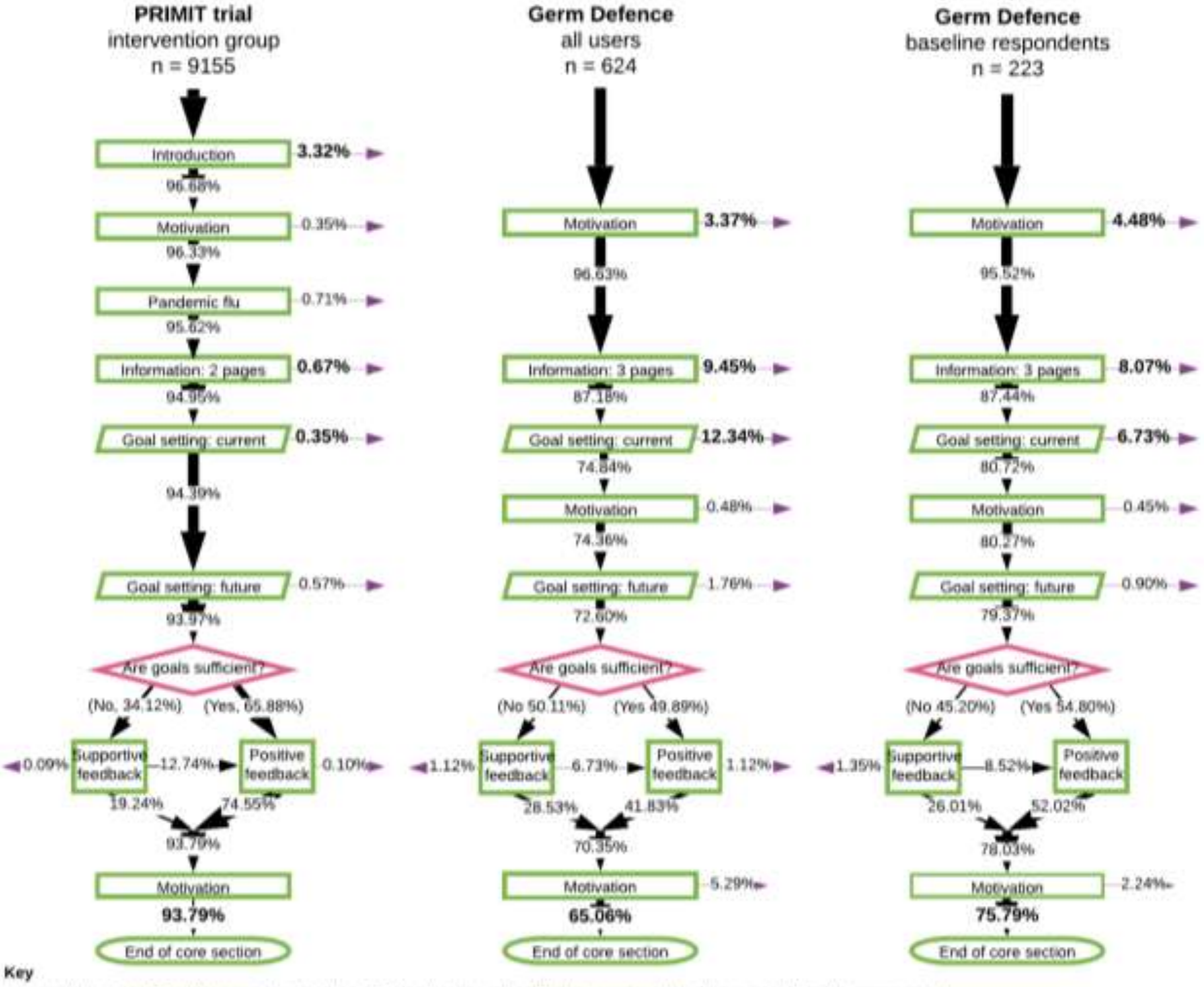
Measure	*N*	*M*	*SD*	Min-Max
Responders				
PRIMIT Study	8959			
Germ Defence	250			
Gender split (female)				
PRIMIT Study	5002/8943 (55.93%)			
Germ Defence	184/243 (75.72%)			
Age (years)				
PRIMIT Study	8945	56.64	13.63	18-94
Germ Defence	250	42.11	13.04	16-74
Current daily handwashing				
PRIMIT Study	8945	3.85	1.15	1-5
Germ Defence	234	3.87	1.09	1-5
Intended daily handwashing				
PRIMIT Study	8944	3.97	1.12	1-5
Germ Defence	231	4.11	1.09	1-5
Perceived likelihood of user becoming ill				
PRIMIT Study	8837	5.11	1.65	1-7
Germ Defence	223	5.52	1.58	1-7
Perceived severity for user				
PRIMIT Study	n/a			
Germ Defence	223	3.61	1.75	1-7

M = mean, SD = standard deviation

When asked how they had heard about the Germ Defence website, 236 users responded (news n=41, 17.4%, word of mouth n=21, 8.9%, social media n=16, 6.8%, search engine n=15, 6.4%, school or children's group n=14, 5.9%, health support group n=4, 1.7%). A further 125 (52.97%) users selected 'other'. When asked to provide further details, 67 (26.80%) users identified themselves as working in healthcare related professions (HCPs). Due to HCPs familiarity with infection prevention behaviours, characteristics for identified HCPs were compared to non-HCPs within the Germ Defence study to examine any potential sample bias. When the sample was split by profession, HCPs indicated higher daily levels of current handwashing (*M*=4.26) than non-HCPs (*M*=3.71)(*MD* = .54, 95% CI [.24, .85]). HCPs also selected higher daily intentions to wash their hands in the future (*M*=4.51) compared to non-HCPs (*M*=3.96)(*MD* = .55, 95% CI [.24, .86]). Although HCPs perceived themselves as more likely to contract an RTI (*M*=6.00) than non-HCPs (*M*=5.33)(*MD* = .67, 95% CI [.21, 1.12]), the perceived likelihood of becoming very ill was similar across all Germ Defence baseline responders (HCPs M=3.25, non-HCPs M=3.62, MD=.04, 95% CI[-.47,.55]).

5.4.2.2 Which parts of the intervention were used?

Attrition was lower in PRIMIT, with fewer users failing to complete the core section (*n*=568, 6.21%) compared to Germ Defence users (*n*=218, 34.94%) (see Figure 10). The pages that saw the most attrition in Germ Defence were those leading up to and including the first page of the goal-setting section, which saw a total of 25.16% (*n*=157) of users leave the intervention, accounting for 72.02% of all attrition. Germ Defence users who had volunteered to complete baseline measures had lower levels of attrition across the all of the core section (*n*=54, 24.21%), and up to the start of the goal setting section (*n*=43, 19.28%). The introduction page to the PRIMIT study saw the highest level of attrition (3.32%, n=304), meaning that 53.52% of total attrition occurred on the first page. Completion of the core section of Germ Defence by users who had identified themselves as HCPs was 79.1% (n=53), which was similar to that of all baseline responders (75.79%).



Key

Diamond: decisions, Parallelogram: data collection, Rectangle: information, Black arrow: direction of movement, Purple arrow: dropout

Figure 10. Flowcharts of percentage of usage for the core sections of PRIMIT and Germ Defence.

After completing the core section, 37.70% (*n*=242) of all Germ Defence users accessed one or more of the three further components, compared to 47.53% (*n*=106) of Germ Defence baseline respondents.

5.4.3 Handwashing behaviours

Germ Defence baseline responders and PRIMIT study users both chose to increase their intended handwashing compared to their current levels for five of the seven situations (see Figure 11, and Appendix J for tables of mean differences and confidence intervals). Intentions to wash hands ‘after going to the toilet’ did not increase in either context (Appendix J, Table 1). Germ Defence users also selected similar current and intended levels for ‘after touching something with germs on’. Scores for HCPs using Germ Defence were not practically different to non-HCPs for both current and intended behaviour (see Appendix J, Table 3). Due to the large sample size, confidence intervals for the PRIMIT study are smaller than those for the Germ Defence sample.

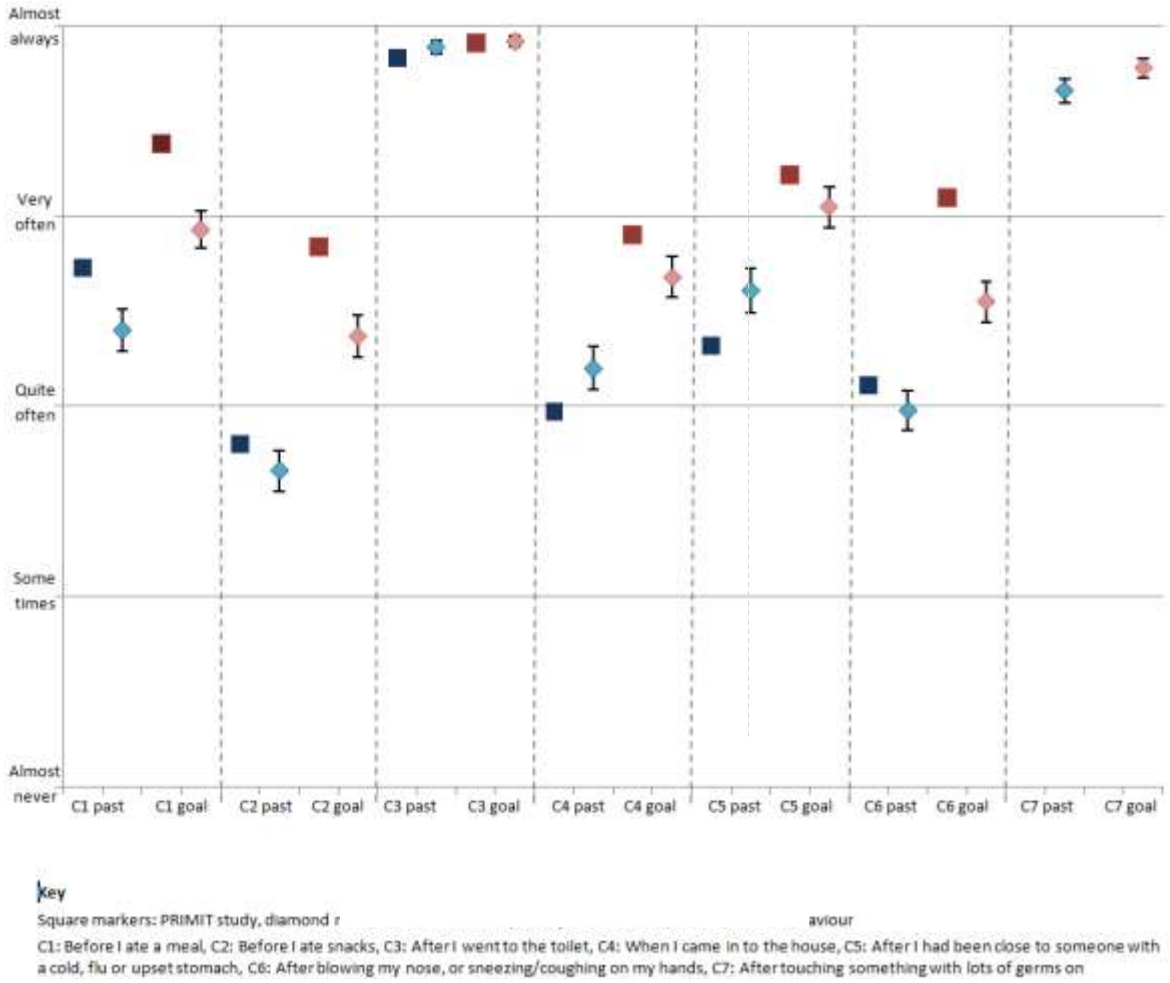


Figure 11. Means and confidence intervals for goal-setting BCT for PRIMIT study and Germ Defence.

When comparing current behaviour between the two studies, scores for ‘before I ate a meal’ was the only practically different situation (*MD* = .32, 95% CI [.12, .52]), with PRIMIT study users washing their hands more frequently (*M*=3.73) than Germ Defence users (*M*=3.41) (see Appendix J, Table 2). Scores for intended behaviour showed that PRIMIT study users had selected higher frequencies (*M*=4.10) for ‘after blowing my nose, sneezing/coughing on my hands’ (*MD* = .38, 95% CI [.21, .56]) compared to Germ Defence users (*M*=3.72).

5.5 Discussion

This paper presents process evaluations of an intervention applied in two different contexts to lower the transmission of viruses in the home by increasing handwashing. The intervention was used in an RCT by NHS patients recruited via their GP (PRIMIT study), and as an open access intervention available to the general public (Germ Defence). By comparing the two contexts we hope to better understand: any effect that the design of the RCT had on who accessed the intervention; how the intervention was used; whether the intervention continued to increase handwashing intentions once disseminated to the public.

5.5.1 Implications for online research enrolment procedures

Half the PRIMIT RCT participants dropped out during the online research enrolment process, prior to reaching the intervention. The intervention had been developed in line with the Person-based approach; accessibility and engagement with content and design were optimized through think-aloud interviews and survey responses (Yardley, Miller, Teasdale, & Little, 2011). However, although baseline measures were assessed quantitatively to ensure they were valid (Miller et al., 2012), additional patient and public involvement and/or qualitative assessment during development would have provided information on the impact and acceptability of these research processes. This would have allowed the opportunity to improve them and reduce the risk of losing a substantial number of potential users before they have even reached the intervention.

The Germ Defence study lost a third of participants during research enrolment. Considerable effort was invested to develop and optimize the consent and baseline measures for Germ Defence. The process was made easier and quicker than for PRIMIT, with optional rather than compulsory baseline measures. These pages were constructed in line with the design values and language used for the original intervention (Miller et al., 2012; Yardley, Miller, Teasdale, & Little, 2011), and were additionally tested for usability by eight researchers experienced in person-based digital intervention development. As a preliminary dissemination study, these procedures were necessary to compare Germ Defence participant characteristics and handwashing behaviours to those from the PRIMIT study. As with the PRIMIT RCT, the potential impact of this on uptake had not been assessed in advance. However, by making the process voluntary we were able to demonstrate that over 60% of Germ Defence users preferred not to complete baseline measures at all.

Almost 80% of first visits to the Germ Defence website did not progress past the first page. The data available do not offer explanation for the high initial drop out. One possible explanation could be that the intervention was first developed in 2008 and intended to be used on personal computers and larger screen tablets; the website was unable to responsively adapt to mobile viewing (e.g. users needed to enlarge or scroll across the page if viewing content on a mobile phone screen). This may have discouraged a considerable number of users. In addition, the number of visits at this point may include multiple visits by the same people, consequently, users who visited the first page and then decided to return at a later point to view the whole intervention cannot be identified within the attrition rate.

5.5.2 Implications for intervention usage

High levels of attrition during research enrolment procedures are a concern as they may lead to selection bias in the remaining sample. Comparison of intervention usage between the two contexts highlighted this effect: over 90% of PRIMIT study users completed the core section of the intervention, compared to 65% of all Germ Defence users. This suggests that effortful enrolment research procedures and high levels of dropout may have resulted in a remaining sample that was more motivated to use the intervention. This is supported within Germ Defence where the completion rate increased to 75% for users who had participated in baseline measures, and they were more likely to view the additional components within the intervention. As none of the user characteristics recorded at baseline (e.g. handwashing behaviour, perceived likelihood of becoming ill) were associated with completing the core section of the intervention, explanations for this increased usage may be explained by alternative characteristics (e.g. motivation to learn about infection prevention behaviours, altruism). As suggested above, extending qualitative methods used to develop intervention content to include research procedures may also help identify characteristics associated with engagement with the research, as well as with target behaviour.

The majority of attrition from the intervention occurred across the first few pages in both studies. Germ Defence users may have chosen to leave as they were dissatisfied with the content, or alternatively, given that these pages received positive feedback during development, users may have felt they had found sufficient information and did not need to continue. Further research is required to understand this attrition, particularly for the introduction page of the PRIMIT study. This page contained no ‘active ingredients’ yet saw the highest proportion of users leave despite having completed the effortful enrolment process. The page with the highest proportion of attrition in Germ Defence was the first page of the goal-setting section. This is the first time users were required to interact with the intervention, and they were unable to progress to the next page without entering current handwashing frequencies. Although users may have decided to leave rather than complete an activity they considered to be too effortful, the goal-setting section was seen to be effective at raising intentions to wash hands more in the future.

5.5.3 Implications for promoting increased handwashing

When comparing PRIMIT study and Germ Defence handwashing selections for the goal-setting BCT, users in both contexts intended to improve their frequency of handwashing in the future in five situations (i.e. before meals, before snacks, after coming in to the home, after being close to someone who is ill, after sneezing or coughing). Minimal improvement was seen for washing hands after going to the toilet and touching something with germs on. This is probably due to the high levels already reported for current behaviour. Interestingly, users reported only washing their hands ‘quite often’ on average after blowing their nose, sneezing or coughing, despite repeated public health campaigns targeting this specific behaviour in the UK (Department of Health and Social Care, UK Government, 2011). Qualitative research during intervention development highlighted that people found this situation difficult to carry out due to lack of control (Morrison & Yardley, 2009; Yardley, Miller, Teasdale, & Little, 2011). However, users did chose to improve on this behaviour in the future having viewed the intervention.

The goal setting section helped Germ Defence users to plan improvements in handwashing in the same way as had been seen in the PRIMIT study RCT for five of the six shared situations. This suggests that the intervention mechanisms work as effectively when disseminated to the general public as they did in the RCT, and that the reduction of illnesses seen in PRIMIT study sample is likely to be replicated in the Germ Defence sample. These similar levels of in intended handwashing occurred despite the RCT experiencing greater levels of selection bias as discussed above, leading to a more motivated sample. An explanation for this may be found in the Germ Defence sample where almost a quarter of baseline responders were HCPs. At baseline, HCPs indicated higher amounts of both current and intended daily handwashing than non-HCPS. Yet when completing the goal-setting section, HCPs and non-HCPs showed similar levels of current and increased intended future handwashing for the specific situations within the BCT. This suggests that although HCPs wash their hands more frequently, they might be over-looking handwashing opportunities in the home that are important for reducing infection transmission. If this effect was seen with HCPs using Germ Defence, then it is likely that using these specific situations was also effective for highly motivated users in the PRIMIT study.

5.5.4 Application of the findings to disseminate Germ Defence for use in a pandemic

In March 2020 Germ Defence was adapted and disseminated globally for use in the COVID-19 pandemic (www.germdefence.org; Ainsworth et al., 2020). The findings from this preliminary dissemination study were applied to the intervention to maximize uptake and reach. Having established that even minimal research enrolment procedures may act as a barrier to usage, all baseline measures were removed for the Germ Defence Covid-19 intervention so that users could access the intervention as quickly and easily as possible (Ainsworth et al., 2020). Instead, after completing the core section users are invited to take part in a survey where some demographic information is collected along with questions about using the intervention. Whilst the sample that reaches this stage of the intervention will be biased, the survey has still collected valuable and insightful information (Morton et al., 2020). In addition, the goal-setting section imbedded within the intervention provides a measure of past/intended behaviour. To reduce attrition within the goal-setting section, an additional message explaining the value of completing the measures has been added, which is hoped will encourage participation.

5.5.5 Limitations

Comparing the large sample in the PRIMIT study to the much smaller number of users who accessed Germ Defence is problematic because they are likely to differ in many respects which could not be evaluated. However, comparisons are often made between implementation of interventions in different contexts, and given the difference in contexts and probably also populations, it is notable that beliefs at baseline and self-reported behaviours were similar in the trial and the implementation.

The difference in number of participants is likely to reflect recruitment methods: the PRIMIT study sent letters directly to over 800,000 people, whereas Germ Defence relied heavily on health organizations to distribute the details. The low level of Germ Defence users suggests that prevention of seasonal colds and flu may not be a strong motivation for many people - Germ Defence for COVID-19 has seen considerably higher usage (Moore et al., 2015). This means that people who did use Germ Defence may have been more highly motivated to practice good infection prevention behaviours, and given the low proportion of uptake for PRIMIT, the same may be said of that sample. Yet despite this, the intervention group for the PRIMIT study experienced less RTIs, and selections for the goal-setting section demonstrated the ability of the intervention to help users in both studies plan to increase their handwashing.

The large proportion of dropout seen on the first page of Germ Defence is unexplained. This is a common phenomenon of intervention usage in the community (Szinay, Jones, Chadborn, Brown, & Naughton, 2020), and as such, including qualitative data collection in future process evaluations would enable some insight for this problem and inform further refinement of the first page to maximize usage.

5.5.6 Conclusions

This study provides an example of how interventions assessed through RCTs can be examined and adapted to try to optimize their usage when disseminated to the general public. By comparing RCT data to a novel, preliminary dissemination study we were able to examine our three aims. First, this research demonstrates that the Germ Defence intervention continues to raise users' intentions to wash their hands more in specific situations where the risk of virus transmission is high, replicating the behaviour changes reported during the PRIMIT RCT. The responses show that using a goal-setting BCT was helpful for increasing intended handwashing, including in situations identified as being particularly difficult despite having been already targeted by long-term public health messages. For the second aim, we established that the effortful online enrolment procedures required for the PRIMIT RCT led to participant dropout, acting as a barrier to accessing health information. The effect of this was apparent through our third aim of comparing intervention usage, as the remaining PRIMIT users showed higher levels of usage than those accessing the intervention in the Germ Defence study. By designing the Germ Defence study to observe and compare the preliminary dissemination of the intervention, we were able to establish points of increased attrition within the intervention. The findings provided the opportunity to adapt and improve Germ Defence for wider dissemination, and during the onset of COVID-19, for a public health emergency.

5.5.7 Acknowledgements

We would like to thank the Economic and Social Research Council for funding this paper (award: 1692136 – Understanding digital intervention engagement: Making sense of large-scale usage data). The PRIMIT study was funded by the Medical Research Council. The intervention used for the PRIMIT and Germ Defence studies was developed using LifeGuide, which was partly funded by the NIHR Biomedical Research Centre (BRC), Southampton. LY is a National Institute for Health Research (NIHR) Senior Investigator and is affiliated to the NIHR Applied Research Collaboration (ARC) West and the NIHR Health Protection Research Unit in Behavioural Science and Evaluation of Interventions at the University of Bristol in partnership with Public Health England (PHE).

5.5.8 Conflicts of Interest

None declared.

Chapter 6 General discussion

This discussion section includes:

- An overview of the research and consideration of how the aim and objectives were met.
- Details on how the AMUsED framework was iteratively developed and usage analyses undertaken, to enable a better understanding of the research process and my role in leading it.
- Consideration of the contribution of the research to the wider field.
- Limitations of the research along with future directions for the field and suggestions for advancing the research.

6.1 Overview of the research

The aim of this research was to develop a methodology to structure the meaningful analysis of large-scale usage data collected from digital behaviour change interventions (DBCIs). The AMUsED framework is a novel and detailed approach that supports systematic usage analyses. By completing the three checklists, researchers are able to identify meaningful ways to operationalize usage and generate research questions with which to structure the analyses, leading to findings that are able to meet the key aim of process evaluations: explaining how an intervention was used, when and by whom (Moore et al., 2015). The framework encourages researchers to consider the context of the data in relation to the intervention structure (e.g. content, theory-base) rather than focusing solely on the datasets themselves. These types of analyses, that are based on meaningful measures of usage, are able to produce findings with insight about effective levels of engagement with digital interventions (Yardley et al., 2016).

Details of the framework formed the first paper in this thesis, but the development of the framework was iterative. At the outset, the research had been intended to support analyses of secondary data sets. However, the two usage analyses reported in the subsequent papers also informed the development of the framework as well as being guided by it. Due to my previous work on the PRIMIT study and interest in antimicrobial resistance (AMR), I selected the PRIMIT study and Internet Dr secondary datasets to analyse. I was also working on adapting the PRIMIT intervention to become Germ Defence, and decided to apply the framework to support this process, thereby extending the checklists to be relevant during intervention development as well. This is important as it encourages researchers to consider in advance what usage data will be collected, the format it is recorded in, and how it may be analysed. Without this process there is an increased risk of finishing a trial with only summative usage metrics available, thereby limiting the scope of possible analyses.

Five specific objectives were identified as necessary for reaching the research aim:

- To explain how large-scale data from digital health interventions can be best utilised to develop our understanding of engagement.
- To develop a method for analysing large-scale data that is systematic and efficient, and produces findings that are meaningful to the intervention and inform our understanding of how theoretically-underpinned interventions are used.
- To enable the consideration of appropriate data analyses and required data collection during the development stage of a digital intervention, and to support post-hoc analyses of pre-collected data.
- To apply and test the methodology to demonstrate how insights may be gained for understanding intervention efficacy, leading to suggestions for improvement, and informing the wider field of engagement with digital interventions.
- To describe concepts currently used to define and analyse usage, and critically discuss how they have been applied to DCBIs, to demonstrate the contribution of the PhD research within a rapidly developing field.

The first two objectives were met through the development of the AMUsED framework and the three checklists reported in Paper 1 (see Chapter 3, Section 3.7). The framework supports researchers in breaking down large-scale usage data and analysing them in a way that is meaningful to the intervention. This also overcomes the challenges identified in measuring usage of DBCIs, such as relying upon summative usage metrics to try and understand fine-grained interactions, and the idea that more usage will lead to better outcomes (see Introduction, Chapter 1). Through the framework, I propose selecting usage metrics based on how meaningful they are to the structure of the intervention, its contents and theory base. The framework also supports generating research questions to structure inferential analyses examining usage and outcomes, and also user characteristics. Through these analyses it is possible to identify what type of engagement was effective and for who. The third checklist in the framework also encourages researchers to consider the format in which their data is collected, so that large-scale data sets are compatible with the analytical tools selected. As highlighted earlier in the Scoping Review, the AMUsED framework is the only method available that is able to meet these objectives (see Chapter 2).

The next two objectives were achieved through the application of the AMUsED framework to structure the process evaluations for the Internet Dr and PRIMIT/Germ Defence papers (see Chapters 4 and 5). For the Internet Dr and PRIMIT study, the framework was applied to shape post-hoc usage analyses. In comparison, for the Germ Defence study the framework was used to help shape data collection during the intervention adaptation and to create an a priori plan of analysis. Both papers demonstrate the value of the framework for leading to findings that provide specific guidance on intervention improvement. The recommendations contribute to our understanding of engagement: improving intervention/study design to reduce dropout and maximise the number of users who reach important behaviour change content; highlighting patterns of engagement that lead to behaviour change; providing insights for future implementation. In addition, by utilising metrics that represent usage of theory-based content it was also possible to examine the effect of psychological theories and behaviour change techniques in outcomes. Both interventions were designed to be used to support NHS primary care services, and the findings also provide clinical insight for the areas of self-caring for minor ailments and infection prevention (see Clinical Implications section later in this chapter).

The final objective drove the Scoping Review reported in Chapter 2. This review of reviews captured the conceptualisations of ‘engagement’ and ‘adherence’ and how they impact upon the way that usage analyses are undertaken. Within this research, intervention usage is considered to be a dimension of engagement along with users’ subjective experience of the intervention (Perski et al., 2017). Intervention interaction is therefore considered to be led by the user and their perceptions of acceptability, expectations and target outcomes.

6.2 Reflections on the research process

As highlighted at the start of the discussion, the process of developing the framework became iterative due to my role in the research and familiarity - or lack of - with the interventions (see Figure 12). This led to challenges and opportunities in developing the framework which were not discussed in the three papers due to word constraints, but are important in understanding the research process and my contribution in leading it.

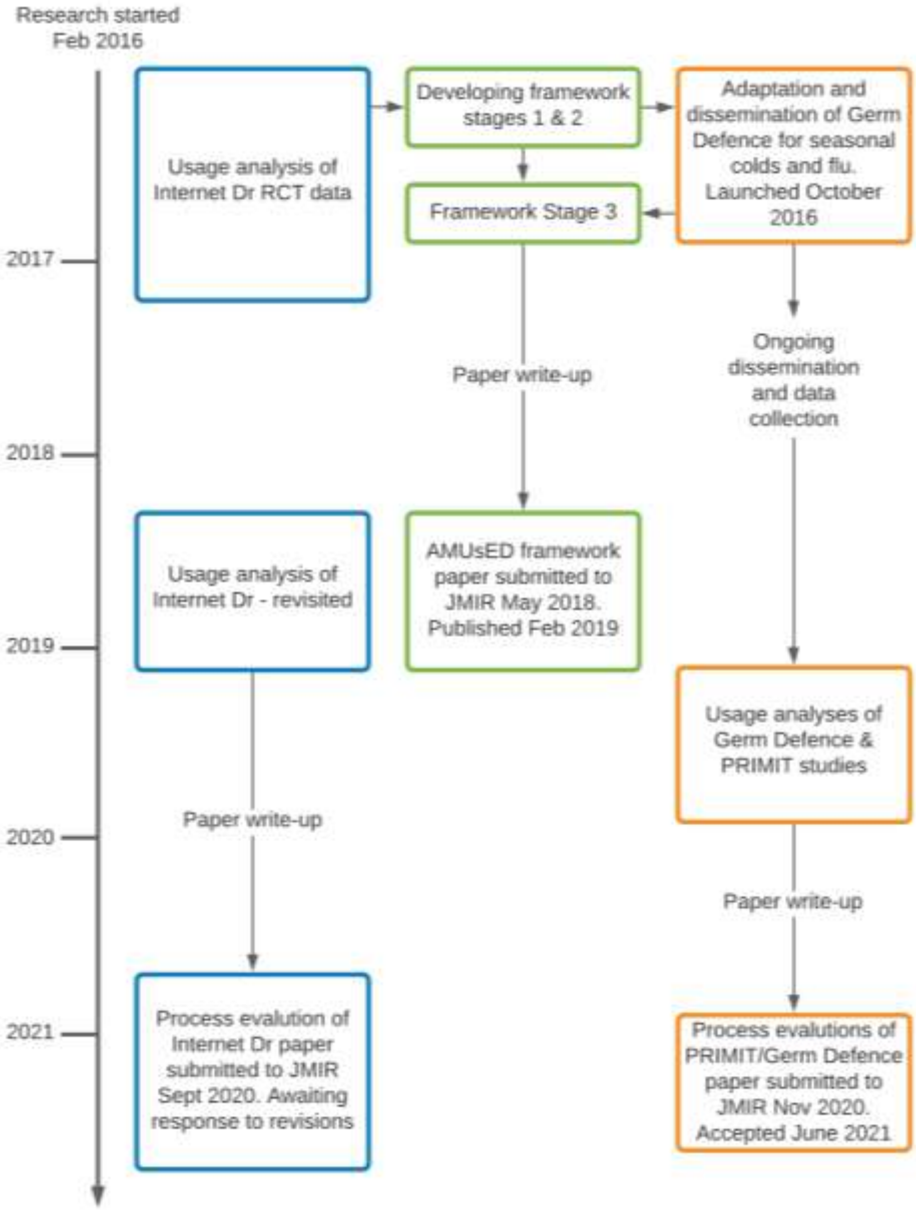


Figure 12. Thesis timeline of the three papers and the interaction between them.

6.2.1 Iteratively developing the AMUsED framework

The Internet Dr RCT showed that using the intervention encouraged users to appropriately self-care for RTIs rather than make unnecessary visits to their GP. My research began with preliminary analyses of the usage and self-report data collected during the RCT with a view to understanding how this change occurred. As I was previously unfamiliar with the intervention, I documented the process with the intention of designing a standardised method for carrying out an analysis which would simplify the process when applied to other interventions. As a starting point, I examined the summative measures of usage most typically reported in usage analyses (e.g. completing the study at 6 months, number of logins, number of pages viewed and time spent on the intervention) for associations with the outcome measure of patient enablement index (PEI) scores. However, the results produced were unable to identify patterns of usage that increased enablement. The next stage was to analyse associations between these same usage metrics and user characteristics, with the intention of identifying whether certain types of users had engaged differently with the intervention. Again, the analyses were unable to provide any insight into who used the intervention based on these summative measures of usage.

Leanne Morrison had previously argued that these types of summative measures of usage often provide only limited understanding for how an intervention is effective as they do not capture patterns of usage specific to an intervention (Yardley et al., 2016). As this was the case for Internet Dr, different usage metrics were needed that would be meaningful to the intervention. To identify these it was necessary to consider the intervention structure. At this point I realised that although I was by now very familiar with the usage data, the architecture of the intervention was still unclear. As a result, I started the process of familiarising myself with the intervention by obtaining the intervention software files containing the web-pages and coding showing the contents of the intervention, how it was delivered and when data was collected. This process revealed that although the intervention content was available at all times throughout the study it was intended to be accessed when a user was experiencing RTI symptoms. This suggested that constructing usage metrics to enable comparison of users who were ill with those who were not, would potentially lead to a clearer understanding of how the intervention had been successful, and for whom.

Documenting this process highlighted that unfamiliarity with an intervention hinders a researcher’s ability to select meaningful usage metrics. As in-depth usage analyses may only be undertaken after data collection, often towards the end of a research project, issues such as staff changes or lack of resources mean that these analyses may frequently be carried out by researchers who were not originally involved in developing the intervention – as was the case with my analysis of Internet Dr. Therefore, the process of familiarisation with both the intervention structure as well as the data is crucial. Through this realisation, I developed the first two stages of the AMUsED framework. Stage 1 encourages researchers to consider not only the data, but to familiarise themselves with the intervention through the content, timing of delivery, data collected and the theoretical underpinning, making it possible to identity all potential usage metrics. Stage 2 focuses on which of these metrics are meaningful to the structure and are likely to provide the greatest insight in to how the intervention was used successfully, recognising that this is likely to differ by intervention. By generating concise research questions around the selected usage metrics (taking in to consideration the theoretical underpinning of the content), inferential analyses can then examine associations between these metrics, outcomes and user characteristics.

The third stage of the framework was developed through my work on adapting the PRIMIT study intervention to Germ Defence. I was keen to extend the remit of the research and compare usage data collected from the publicly disseminated intervention to the RCT data as this would be a good opportunity to test the applicability and contribution of the framework to other interventions during the development phase as well as for post-hoc analyses. This would also enable evaluation of whether the intervention was still effective when used ‘in the wild’, which was an important next step in the implementation process. The first two checklists were applied to the PRIMIT study and further adapted to make them more inclusive for different types of interventions. Whilst considering the data and appropriate research questions, it was apparent that some of the research questions were unanswerable with the PRIMIT data. For example, users who selected low or no increase in handwashing behaviour in the goal-setting section were encouraged to repeat the exercise, yet the data only contained one completion for each user as the new selections had over-written the previous ones. This led to changes in how the data were collected and stored for Germ Defence. Through this process, it was evident that the framework was also helpful for considering data collection in advance, and that a third checklist would also be useful to prompt consideration of data structure and format for compatibility with different analytical tools.

The research supervisory team had originally envisaged that the first paper of this thesis would report the findings from the Internet Dr process evaluation along with a description of the methodology used to structure the analysis. However, I believed the framework had developed to an extent whereby combining both the Internet Dr findings and the methodology would not provide enough detail on either, and that the framework was substantial enough to justify a purely methodological paper with Germ Defence and Internet Dr included as case studies.

6.2.2 Undertaking usage analyses

The research questions generated for Internet Dr and reported as a case study in the first paper, focused on analysing usage based on whether participants were ill or not at the time of accessing the intervention. However, once the analyses started it became apparent that very few users had accessed the intervention when ill (n = 59, 3.96%). This meant that further analyses of this sub-group would be insufficiently powered, but more importantly, this proportion of users was unlikely to account for the level of behaviour change seen in the intervention group. On that basis, a different metric was required to capture meaningful usage and the research questions needed revising in line with the framework checklists. Having adapted the checklists to apply to PRIMIT/Germ Defence, when revisiting them they were now clearer with more detail and structure. Through the process of completing the first checklist again, flowcharts capturing page flow, contents and points of data collection were created. From these it was evident that only the pages containing advice on self-caring for RTI symptoms were intended to change behaviour (see Figure 4, Chapter 4). This process of familiarisation was illuminating – not only for the new information provided on selecting meaningful usage metrics, but also in support of the value of the framework for guiding the process.

Viewing advice pages was identified as a meaningful usage metric and new research questions were generated to examine both the types of advice pages viewed and the context in which they were accessed. Inferential analyses found that viewing advice pages constituted effective engagement and people who had previously visited their GP for an RTI were more likely to view them. The LifeGuide Visualisation Tool (LVT) was used for sequential data analysis to examine when content was accessed. This revealed that the majority of users accessed advice pages immediately after having enrolled for the study – in direct contrast to anticipated usage whilst experiencing an RTI, and providing important implications for future implementation.

The Internet Dr usage analysis demonstrates the flexibility of the framework to be applied iteratively and revisit checklists if unanticipated results are revealed, as described in the framework model (Figure 3, Chapter 3). In addition to the findings reported in the second paper, the framework also helped to identify further potential analyses (e.g. comparing usage for different illness symptoms, identifying the order in which components were viewed in, repeated use of advice pages within components). However, the reported results were limited by the size of the paper.

The Germ Defence usage analysis was relatively straightforward and easy having applied the framework during the development phase and created an analysis plan in advance. The PRIMIT and Germ Defence research questions reported for the case study in Paper 1 remained unchanged. However, the PRIMIT study usage analysis was considerably more difficult due to issues with the data such as: not being recorded in an easily accessible format; specific variables being complicated to extract (e.g. which type of feedback had been viewed for the goal-setting section); partially incomplete collection (e.g. usage of research procedures); over-writing previous data. Had the framework been available during the development of PRIMIT, I have no doubt that these issues would have been identified and the analysis would have been considerably easier and quicker.

As part of the process evaluations, usage of the study enrolment procedures (e.g. consent, baseline data collection) and intervention sections were examined separately, thereby distinguishing between engagement with the intervention and engagement with the research. Research enrolment procedures are necessary to ensure ethical standards are met, yet they may be considered unwieldy and effortful by participants especially when compared to procedures used for commercial interventions. This may result in increased attrition and sample bias. One of the aims in designing the Germ Defence study was to capture and quantify this effect, leading to recommendations for the design of research procedures in the future. As with the Internet Dr, the framework checklists also identified considerably more usage metrics than were reported for the paper (e.g. whether optional pages were accessed, the number of users who printed out the handwashing plan, users who reviewed the plan). However, when working on the COVID-19 version of the intervention, I applied the findings from these additional analyses, along with those related to the research procedures, to make Germ Defence as accessible and engaging as possible. For example, the PRIMIT/Germ Defence intervention had optional pages available at the start containing research details that were intended to increase perceived credibility and trust in the website. Yet the usage study highlighted that accessing these pages increased attrition. As a result, they were subsequently placed in an optional component after the core section for the COVID-19 version.

6.3 Contribution of the research

6.3.1 Providing a practical, applicable tool

The AMUsED framework contributes a useful method for analysing usage data in way that provides meaningful insights about effective engagement with digital interventions. Completion of the first framework checklist may initially appear time consuming as it requires the researcher to immerse themselves in the intervention contents and structure. However, by identifying a limited number of meaningful usage metrics and then generating research questions using those variables, the framework presents a more efficient way of working. In addition, identifying research questions in advance using theory- and content-based rationales avoids the potential to data-dredge or find results that do not provide useful insight about how the intervention was used effectively. Applying the framework during intervention development encourages thoughtful and optimised data-collection that will lead to richer data and ensure the required variables are recorded. Considering in advance the format in which the data will be collected for compatibility with prospective analytical tools will also save considerable time during the analysis.

To date, the AMUsED framework has been cited in fourteen published research articles and a book. The majority of these acknowledge the framework’s contribution for providing guidelines for usage analyses (Ensari & Elhadad, 2021; Issom et al., 2020; Nelson et al., 2020; Stragier et al., 2019; Xie, Itzkovitz, Roy-Fleming, Da Costa, & Brazeau, 2020). Pham, Shaw et al. (2019, p.2) also recognise the framework’s role in the wider field of research: “*Scholars have also pushed for new theoretical frameworks to guide more systematic log data analyses and support transparent and replicable evidence generation* (Miller et al., 2019a). *The aggregation of this research productivity* *has advanced the scale and spread of research analytics in industry and academic evaluative practice*". The framework has been cited for providing the definition and ability to identify meaningful usage metrics (Harjumaa et al., 2020; Morton et al., 2021; Nurmi et al., 2020; Perski, Lumsden, et al., 2019), and the importance of this: "*Thus, as mHealth apps proliferate, measuring what might be considered meaningful or effective engagement, that is, engagement that leads to a desired outcome, is necessary*" (Alshurafa et al., 2018, p.2).

Two studies applied the framework to structure their usage analyses (Shi et al., 2021; Stassen, Grieben, Froböse, & Schaller, 2020). Shi et al. (2021, p.8) reported each of the three stages in relation to their intervention and found that applying the framework "*can enable more useful analysis but also improved decision-making during the design and optimization process*". In another study, the researchers applied the framework retrospectively to understand why their usage analysis had not produced as insightful findings as they had hoped (Dunphy et al., 2021). They concluded the study had been "*limited by the specificity of usage measurement that TRAK-ACL was capable of when considered against the AMUsED framework, a standard for reporting usage in digital trials. In a future trial, TRAK-ACL would need more specific usage measurement capabilities*" (Dunphy et al., 2021, p.8). These studies are by researchers outside of the University of Southampton Digital Intervention Group, the interventions were developed using different software and address a range of health issues. This suggests that the framework is flexible and adaptable enough to be suitable for a variety of interventions, and also supports the need for a systematic methodology in the field.

6.3.2 Compatible with qualitative development and evaluation methods

The framework can be incorporated in to mixed-method intervention development approaches, and has been used as part of the person-based approach (PBA; Yardley, Morrison, Bradbury, & Muller, 2015). The PBA focuses on qualitative methods to supply rich, insightful information that helps develop engaging and useful interventions. However, this type of information may be less informative for highlighting whether negative feedback translates in to disengagement when used, and for pin-pointing exactly where that may occur. For example, digital interventions typically experience at least 25% dropout on the first page, whereas the proportion of participants who would indicate, or even enact, this level of disengagement during an interview is most likely substantially lower.

Incorporating the framework to explore quantitative usage data is compatible with qualitative methods, so that the findings from each are additive and better able to explain users' experiences (Moore et al., 2015). For example, in the Internet Dr process evaluation a third of users completed the enrolment and baseline measures but then left when they reached the first page of the intervention. We can speculate that this was because they were not experiencing an RTI at that time and had intended to return if they became ill during the trial period. However, qualitative data are the only way to ascertain whether this was the case. Had this been established, it would be simple to add a message to the first page explaining that the intervention is effective if used when well and encouraging users to continue. Conversely, the high level of attrition on this page wasn't identified during the qualitative interviews carried out for the development of Internet Dr, and provides a good contextual example of when disengagement may be highlighted through usage data.

6.3.3 Designing engaging and effective interventions

Effective engagement has been described as the minimum threshold of usage necessary to bring about positive outcomes (Yardley et al., 2016), and provides a useful lens through which to design and evaluate interventions. By identifying levels, types or patterns of effective engagement, it is possible to provide practical insights in to how an intervention may be improved by ensuring that more people reach this threshold. This will lead to interventions that are streamlined and less effortful for the user, and more likely to meet users' requirements and expectations of acceptable usage. This in turn may produce higher levels of efficacy through reduced dropout by users who feel the intervention requires too much input or is not relevant.

To identify effective engagement it is necessary to consider different aspects of usage including both summative (e.g. viewing the first session of the PRIMT study intervention) and fine-grained measures (e.g. viewing advice pages for Internet Dr). Reliance on purely summative metrics to explore patterns of usage makes identifying effective engagement more challenging, and is less likely to capture contextual effects or differences within specific user groups. However, applying the framework's key concept of meaningful usage increases the likelihood of establishing which type of usage leads to effective engagement. For example, having identified viewing advice pages as effective engagement through the Internet Dr usage analysis, further examination of the fine-grained details for when they were viewed and by who, revealed more in-depth information on how the intervention had been effective.

The framework methodology can also be used to analyse data from interventions that did not achieve their desired outcomes. Whilst primary analyses tend to focus on change across the whole sample, process evaluations are able to explore the finer details of who the intervention may have been successful for and what type of usage had a positive effect. Understanding how, where and why interventions *do not* meet expectations provides as great a contribution to the field as process evaluations of successful interventions (Craig et al., 2008). For example, although the Internet Dr RCT demonstrated that the intervention is effective at lowering unnecessary GP appointments, the process evaluation identified points of attrition where substantial proportions of users were lost: the intervention welcome page and during the Doctor’s Questions component. This information provides the opportunity to improve the intervention by reducing the number of compulsory pages prior to reaching advice pages, and adding information explaining that the intervention is still effective for users who are not experiencing an RTI. Making these relatively simple changes is likely to broaden the efficacy of the intervention by increasing the number of users who reach advice pages.

6.3.4 Examining current theory

Although this research was not intended to develop a new theory or directly advance an existing one, application of the framework methodology is able to contribute to our understanding of psychological theories when utilised to change behaviour. Analysing usage data with specific usage metrics that relate to theory-based sections of content and BCTs is more likely to provide findings that broaden our understanding of the context and types of health issues for which these are effective. For example, when structured by the AMUsED framework, the Internet Dr process evaluation included analyses that tested the intervention logic model. By selecting viewing advice pages as the meaningful usage metric it was possible to compare usage of the three theory-based components. This established that accessing advice pages from any component was effective, with using the Treatment Options advice pages associated with the highest increase in enablement. Content in the Treatment Options component is based on Social Cognitive Theory (Bandura, 1997) and is intended to help users experiencing distress about their symptoms. These findings suggest that this approach is helpful in lowering unnecessary GP visits for minor ailments. This is also the case for the PRIMIT and Germ Defence process evaluations where analysis of the goal-setting section evidenced the effectiveness of this BCT for increasing handwashing frequency across several situations.

The aims of a process evaluation are to assess: implementation processes; mechanisms associated with changes in outcomes; contextual factors associated with variations in outcomes (Craig et al., 2008). In-depth usage analyses, as supported by the AMUsED framework, are better able to meet these aims compared to the descriptive analyses of summative usage metrics relied upon in many studies (see Scoping Review, Chapter 2). With regard to identifying mechanisms of change, behavioural determinants captured at baseline and outcome are frequently analysed to better understand their role in behaviour change. However, they are less examined for their role in mediating the relationship between usage and outcomes, as encouraged in the framework. In the case of Internet Dr, these analyses showed that lowered reliance on HCPs over the course of the trial did not mediate the relationship between viewing advice pages and increased enablement to self-care for RTIs. Although this means that the underlying mechanisms of change for the intervention were not explained by the behavioural determinants selected for the trial, by examining the interaction between usage, outcomes and changes in behavioural determinants, the findings eliminate reliance on HCPs and provide insight for other potential barriers to self-caring that may be considered in the future (e.g. cost of un-prescribed medication, perceived illness severity; Richardson et al., 2018).

Theories of engagement with digital interventions are still relatively new and evolving, with the terms ‘adherence’ and ‘engagement’ frequently used interchangeably (see Scoping Review, Chapter 2). However, this research clarifies these concepts and considers their implications on how to shape intervention design and evaluation (see Introduction, Chapter 1). Through application of the framework, it is possible to understand the types and patterns of engagement that are effective in bringing about positive outcomes. This in turn will help explain why the concept of medical adherence is not necessarily transferable to DBCIs. Instead, identifying realistic, acceptable and meaningful levels of usage that are effective will aid the design of interventions within those parameters.

6.3.5 Clinical implications

In addition to the contribution of the process evaluation findings to our understanding of engagement and psychological theories, they also provide insight for the targeted health issues. The findings from the PRIMIT/Germ Defence process evaluations highlight the importance of promoting handwashing in situations with increased potential for virus transmission (e.g. before eating). By analysing the self-report measures of current and intended behaviour, it was apparent that the goal-setting section enabled users to increase their handwashing, even for HCPs who already reported higher daily levels of handwashing. The lowest frequencies for current behaviour were reported for the following situations: before eating snacks, after coming in to the house, after being close to someone with RTI symptoms, after blowing my nose, sneezing/coughing on hands. This suggests that users had not previously considered these situations as higher risk for transmitting viruses. By explaining transmission routes and providing a variety of cues, users were able to understand how they might infect themselves and also the people they live with. The comparisons between the goal-setting measures in the randomised controlled trial and those from the publically available version suggest that despite the differences in uptake and context, the intervention is as effective at lowering RTI occurrence and onward transmission when used in the wild as it was during the RCT.

The findings from the Internet Dr process evaluation suggest that the intervention design may be effective for encouraging self-care of other minor ailments. As highlighted in the paper, these type of illnesses lead to a substantial drain on NHS primary care through avoidable GP appointments, whilst causing inconvenience and potential distress to the patients. Viewing advice from the Internet Dr raised users' enablement to self-care, even if they did not have an RTI at that time. As most people will have previously experienced RTI symptoms (e.g. temperature, cough, sore throat), it is likely they were able to draw upon this experience and use it when seeking advice from the intervention. In addition, demonstrating that users who had visited their GP for an RTI in the year prior to trial were more likely to view advice pages suggests that the target population were reached and that these users felt that the DBCI was a suitable replacement for a GP consultation in this situation.

Both the Internet Dr and Germ Defence process evaluations contribute to wider evidence and research focused on antimicrobial resistance (AMR). Reducing unnecessary antibiotic prescribing is a primary focus in the fight against AMR, and both interventions were shown in their RCTs to reduce GP visits, with the intervention groups having less prescribed antibiotics than the control groups. The recommendations for improving the interventions identified in the process evaluations will lead to more people accessing and using them, hence, lowering the burden on primary care and the risk of prescriptions for unneeded antibiotics.

6.3.6 Adapting Germ Defence for COVID-19

The timing of the process evaluations for the PRIMIT study and Germ Defence intervention was opportune, with the analysis completed and writing of the paper in progress when the outbreak of COVID-19 started. As the intervention had originally been designed for use specifically in a pandemic, the wider research team was awarded rapid response funding by UKRI to be optimised and disseminated specifically for COVID-19. The findings from my process evaluations were fundamental for maximising reach and inclusion of this adapted version of Germ Defence. The three checklists from the AMUsED framework were applied to structure the COVID-19 intervention and to make advance process evaluation plans.

The PRIMIT/Germ Defence process evaluations paper now represents part of a series of studies for developing, trialling and implementing a digital intervention: acceptability (Morrison & Yardley, 2009), content development (Miller et al., 2012; Yardley, Miller, Teasdale & Little, 2011), exploratory trial (Yardley, Miller, Schlotz & Little, 2011), randomised controlled trial (Gold et al., 2021; Little et al., 2015), behavioural analysis (Ainsworth et al., 2016), preliminary dissemination (Paper 3, in print), rapid optimisation (Morton et al., 2021) and dissemination for a health emergency (Ainsworth et al., 2020; Horwood et al., 2021).

6.4 Limitations and future directions

6.4.1 Technological advances

Digital technology is constantly changing, upgrading and expanding. It is anticipated that the key concepts underpinning the framework (i.e. identifying meaningful usage, generating appropriate research questions) will remain pertinent. However, the checklists within the framework may well need revision or addition to stay in line with new technology. At present the framework does not include macro engagement (i.e. behaviour away from the intervention). As intervention delivery and measurement devices become more portable, and the availability of just-in-time adaptive interventions (JITAI) increases, the difference between macro and micro engagement will decrease as interventions deliver support when it is most needed (Hardeman, Houghton, Lane, Jones, & Naughton, 2019). This type of data will be important for informing our understanding of effective engagement. However, when using JITAI there is the potential to create larger and more unwieldy data sets - making the application of the framework even more of an imperative. Generating research questions and identifying which usage metrics are meaningful, and how and when they should be collected and stored, during the development phase of the intervention will lead to more compact data sets that are easier to manage and analyse.

6.4.2 Novel statistical methods

Stage 3 of the framework focuses on data collection, with the underlying assumption that novel statistical methods and analytical tools may be utilised, but does not consider appropriate types of analyses. This was intentional as this topic is too vast to be adequately addressed, and outside of the aims of the PhD and resulting framework. However, the development of new statistical methods is a significant characteristic of the field of usage analyses, and has the potential to influence the type of findings they are able to generate. Methods such as cluster analyses (Bricker et al., 2018), ecological momentary assessments (Potts et al., 2020), growth mixture modelling (Coa, Wiseman, Higgins, & Augustson, 2019), sequential data analyses (Arden-Close et al., 2015; Morrison & Doherty, 2014) and time series analyses (Beard, West, Michie, & Brown, 2016) evaluate the data by grouping users based on similar characteristics or engagement, identifying usage patterns, and capturing behaviour across time. These analytical tools provide the opportunity to examine data in more detail and complexity. For example, the LifeGuide Visualisation Tool (LVT) was used for the Internet Dr usage analysis to carry out sequential data analyses. It was through this output that usage when well was identified as being effective – an unanticipated contextual factor of the findings with practical application for dissemination.

As with traditional statistical methods, the findings are only useful if they are able to describe the essential elements of how the intervention worked, for whom and in what context, or add to current theories and understanding of engagement. A good example of the latter point is provided in a study combining ecological momentary assessments with n-of-1 methods by Perski, Naughton et al. (2019). Through examining individual behaviour, the authors established different types of usage patterns: some users preferred to use the intervention for short amounts each day (e.g. to complete daily 'dashboard' behaviour measures); others accessed the intervention infrequently but when they engaged with a single component they used it for much longer (e.g. a gaming section). Acknowledging that users vary in how they wish to engage, and with what is effective for them, may lead to a body of research that begins to categorise user types, leading to interventions that may be tailored to their needs.

6.4.3 The development of engagement scales

The idea that engagement encompasses how someone uses an intervention (behavioural engagement) and how they feel about the intervention (experiential engagement) (Perski et al., 2017) has led to the development of scales that try to capture both of these aspects. To date, the three most prominent are: the Engagement Index (Taki et al., 2017); the DBCI Engagement Scale (Perski et al., 2020; Perski, Lumsden, et al., 2019); the Twente Engagement with Ehealth Technologies Scale (TWEETS; Kelders, Kip, & Greeff, 2020). Each of these scales intends to capture and categorise how engaging an intervention is through a combination of self-report scales and usage metrics, with the aim of being able to predict future engagement or behaviour change/outcomes.

Whilst comparing, contrasting and reviewing each of these scales is beyond the scope of this thesis, it is relevant to consider how usage is measured and incorporated within them. The Engagement Index (Taki et al., 2017) and the DBCI Engagement Scale (Perski et al., 2020; Perski, Lumsden et al., 2019) include two or more measures of actual usage. However, these are summative measures that are presented with limited or no theoretical rationale for their selection, and no discussion of how these measures improve our understanding of engagement in general or for the specific intervention, in comparison to other available metrics. At this early stage in their development and application, the contribution of these scales is still unclear. Perski et al.(2020) propose that the DBCI Engagement Scale will be able to replace lengthy and complicated usage analyses (Perski et al., 2020; Perski, Lumsden, et al., 2019), while Taki et al. (2017) suggest that their scale will enable better cross-intervention evaluation and inform future changes to interventions. Yet the results generated by these scales do not produce the same level of insight and specific recommendations demonstrated in the process evaluations in Papers 2 and 3 (i.e. identifying effective engagement, providing specific recommendations for improving the intervention, implications for dissemination). In addition, by not examining theoretically-underpinned content, the impact of behavioural and psychological theories remains untested. In contrast, Kelders et al. (2020) constructed the TWEETS with the aim of examining the relationship between the reasons for wanting to use a DBCI (i.e. affective and experiential), how much an intervention is used, and positive outcomes. However, as subjective self-report measures of usage were analysed instead of actual usage metrics, their findings cannot be said to meet this aim.

Engagement scales are appealing for their simplicity and may provide insight for the concept of engagement overall. In addition, by finding that self-report usage was not associated with actual usage, the DBCI Engagement Scale raises some interesting and potentially important questions around the differences between actual and subjective usage (Perski, Lumsden, et al., 2019). However, at present these scales are best positioned as an adjunct, rather than a replacement, to the type of usage analyses presented in this research that aim to provide rich, detailed evaluations.

6.4.4 Comparing interventions

The AMUsED framework paper presents the Internet Dr and Germ Defence interventions as case studies, followed by a section supposedly demonstrating how the framework enables comparison of the interventions. However, this section provides limited insight. At the time of writing the paper, the objective had been to recognise calls by researchers conducting reviews to report standardised, summative usage metrics that are easily categorised and compared, and are perceived to provide findings that would better inform development teams (see Scoping Review, Chapter 2). As the research progressed, the importance of identifying meaningful usage metrics that are specific to an intervention was highlighted by the findings of the process evaluations. This level of specificity is not compatible with the recommendations made for standardised reporting of usage variables, yet the level of detail does provide research teams with useful information on potential structures and content, and relevant BCTs and psychological theories, to use in their intervention.

Once a larger body of process evaluations has accumulated that contains the type of findings promoted by the framework, it is hoped that reviews will be able to collate themes and similarities in patterns of effective engagement. For example, the lack of insightful findings from the comparison section in the AMUsED framework paper is due mainly to the fact that this was written before the usage analyses had been carried out. With both now complete, the importance of accessing behaviour change techniques and/or theory-based content is evident as a common finding. A broad understanding of usage metrics within the context of interventions would provide an important resource for research teams developing interventions (Pham, Graham, et al., 2019).

6.4.5 Applying the framework

The two interventions reported in Papers 2 and 3 are relatively small with limited usage pathways, making the process of identifying meaningful usage fairly simple. Larger, more complicated interventions are likely to have several meaningful usage metrics, potentially making the framework harder to apply and requiring more inferential analyses to identify effective engagement. Despite the Internet Dr and Germ Defence interventions not being complicated, their usage analyses were still iterative, requiring repetition and revision of stages 1 and 2. In addition, how the framework is applied differs depending on whether it is used during the development phase or for post-hoc analyses (see Figure 3, Chapter 3).

The framework has been presented as an evaluation method as part of PBA workshops. When working through the checklists in these sessions, researchers have reported finding them to be simple and intuitive to complete. In contrast, whilst researchers using them unassisted have also found filling out the checklists to be relatively easy, translating them in to usage metrics and statistical analyses can be more challenging, particularly for researchers who are less experienced in this area. Ideas for making the framework accessible as an online tool with more detailed instructions and examples are being considered. As the framework becomes more utilised, future adaptations and iterations are likely to be necessary for application to complex interventions. Creating an online resource will allow these changes to be incorporated and updated versions of the framework readily available.

6.5 Conclusion

Digital behaviour change interventions have the ability to provide health support to users in a convenient and tailored way. These interventions are able to collect in-depth datasets showing how users interacted with the intervention and whether that led to positive outcomes. Through analyses of this data we can broaden our understanding of how interventions are successful, for whom and under what circumstances, leading to a better understanding of applying psychological theories within interventions and improvements in their design. However, these analyses can be time-consuming and effortful, and rely on the necessary data having been captured. This research developed a method for supporting systematic usage analyses that can produce findings that are specific to the intervention and also provide wider insight in to engagement with DBCIs.

The AMUsED framework was designed to help researchers plan for advance data collection and conduct usage analyses that evaluate the intervention implementation process. Through the three stages of the framework researchers are encouraged to: consider the structure, content and theoretical basis of the intervention; select meaningful measures of usage that will help examine these; generate research questions that will provide insight in to how the intervention was effective; collect and format the data to be compatible with relevant analytical tools. Applying the framework to structure three process evaluations of two interventions demonstrated the value of the framework for producing findings that are practical and applicable for improving the intervention and its implementation. The process evaluations also contribute to our understanding of engagement by identifying patterns of effective engagement and the role of psychological theories in changing behaviours. In addition, the findings have clinical relevance for each of the target health issues and for the wider field of AMR through increased understanding for improving infection prevention behaviours and how to enable patients to self-care for RTIs.

The potential for process evaluations to inform the implementation of interventions for use by the general public, identify mechanisms of changes and contextual factors associated with variations in outcomes is increasingly recognised (Craig et al., 2008). To produce this level of in-depth findings it is crucial that usage analyses move away from merely describing summative measures of usage, and instead to carry out inferential analyses of meaningful usage metrics that are guided by specific research questions examining how the intervention was used, when, and by who. The methods and findings from this research have demonstrated their value through application to the live version of the Germ Defence intervention for COVID-19, where they were used, alongside qualitative methods, to help rapidly optimise content in line with the development of the pandemic. Widespread application of the AMUsED framework would contribute towards a better understanding of how some forms of usage are more effective within the context of an intervention. Through this knowledge it is possible to design digital interventions that are more acceptable, usable and effective for the people who are seeking support.

Appendix A Defining and measuring usage of digital behaviour change interventions: scoping review search protocol

Research questions

- What concepts are used to define usage of DBCIs?
- How are these concepts measured?

Methods

Data bases. Psychinfo, Web of Science, Medline

Search terms. Although this thesis is predominantly concerned with usage, given that many researchers use the term 'engagement' interchangeably with usage, this was included within the search criteria.

Term	Location
Engagement OR usage OR Data OR Attrition OR Adherence	Title
Review	Title
Digital OR internet OR online OR web* OR mhealth OR app OR ehealth	Topic
Health	Topic
Intervention	Topic

Inclusion and exclusion criteria

Inclusion	Exclusion
Digital health behaviour change interventions	• Non health interventions (e.g. education) • Non behaviour change • Non digital
Peer-reviewed papers published in journals	Grey literature (e.g. theses, books, conferences, reports, author's reply)
Reviews (e.g. systematic, scoping, literature)	Non-reviews
1st Jan 2000 – 1st October 2020	Outside of specified dates
Addresses measures /definitions of usage metrics, using the automatically collected log-data recording interactions between the user	Non-usage: • Reports only outcomes/ intervention efficacy • Reports different type of engagement (e.g.

and the intervention.	with services, treatments, stakeholder, community etc…) • Reports only precursors to using the intervention (i.e. app download/adoption, delivery, uptake) or dropout at follow-up • Reports only affective engagement (e.g. satisfaction, usability, acceptability etc…) • Reports only specific intervention components that increased engagement
Includes or applicable to large, in-depth datasets.	• Non app or web-based interventions (e.g. social media, videos, online surveys, patient portals, email and phone text).
English language	Non English language

Data-charting

Form for each paper:

- Author/Year of publication/journal/number of citations
- Type of paper (review, study, commentary)
- Targeted behaviour/health condition/population
- Name of scale, framework, theory, definitions used
- Findings in relation to engagement/usage

The data will be summarised using the narrative review method and material synthesised to identify the order in which the papers were published, their influence on the field and emerging concepts.

Appendix B AMUsED framework stage 1 checklist for PRIMIT study and Internet Dr

Familiarisation with the data – identifying variables		
Generic questions by data type	Intervention: *PRIMIT*	Intervention: *Internet Dr*
1. Intervention characteristics. Data for intervention architecture and content.		
1.1 Workflow. Intervention structure and expected participant interaction and navigation through the intervention.		
How many logins/sessions are available?	*4 sessions. For dissemination: change to 1 session (see previous findings 3.2)*	*Not structured as sessions, access intended during an RTI.*
When are they available?	*1st at baseline, 2nd released 3.5 days post-baseline, 3rd at 10.5 days, 4th at 17.5 days post-baseline.*	*All content available at every access. Prompted logins at: baseline, 5 x 4-weekly interim questionnaires, & follow-up questionnaires. Then during illness and 48 hour follow-up.*
Are new sessions released depending on time elapsed or task-completion?	*Released after the times set, and dependent on having accessed previous session.*	*No.*
Are there limitations on the availability of the intervention?	*Previous session must be accessed in order to view next. Content from previous sessions could be revisited in subsequent session.*	*No.*
Is the purpose of a session to collect self-report measures and/or use the intervention?	*Session 1 includes baseline measures and content, all other sessions are content alone.*	*Questions are asked after login to ascertain user's purpose – either to use intervention for illness or to complete interim questionnaires.*
When is the intervention considered to be finished?	*At the end of the 4th session.*	*The study finishes after the follow-up questionnaire at 24 weeks. The intervention was finished when users no longer wished to access it.*
What prompts are used to encourage usage (e.g. emails, texts, notifications) and when are they sent?	*An email is sent when next session is ready, followed by 2 further emails if they don't login. For dissemination: 2 emails prompting login to follow-up survey.*	*Email prompts for 4 weekly data collection. Additional emails are sent after logging in during illness to prompt repeat login after 48 hours.*

Does the intervention contain 'tunneled' (compulsory) sequences of pages which users have to view to move forward?	*Initial pages of each session are tunneled. For dissemination: tunneled pages from session 1 will be at start of intervention.*	*'Doctor's Questions' component is tunneled as symptoms are assessed before providing illness management advice.*
Are users able to select linked components they wish to view, and avoid others?	*Optional menu components are available after completing tunneled pages. For dissemination: 3 menu components will be available using content from sessions 2-4.*	*There are 3 linked menu components available from the home page.*
1.2. Content. Content available within the pages of the intervention.		
What are the components available?	*Tunneled components in sessions 1-4 providing information on handwashing. Menu components: more advice on looking after someone with flu, details about the research, revisiting goal-setting, revisiting information from previous sessions. For dissemination: tunneled component from session 1. Menu components for more advice on handwashing, flu, and details about the research. There is also the opportunity to return to the start of the tunneled pages again.*	*'Doctor's Questions'. Asks about symptoms & provides advice recommending either: self-management, phoning NHS Direct, or seeking immediate medical attention. 'Treatment Options'. Advice on coping with symptoms: without medication, with medication, and boosting immune system. 'Common Questions'. 2 sets of FAQs: 'Ask the Internet Dr' - medical questions about illness and treatment, 'Common myths about Colds and Flu' - general questions about illness beliefs.*
What is the aim of each component and are they based on underlying theoretical constructs?	*Messages based on the theory of planned behaviour promote handwashing as an effective behaviour (positive attitudes), socially desirable (subjective norms), and easy to do (perceived behavioral control). Messages utilizing protection motivation theory provided information on health consequences and infection transmission for RTIs.*	*Doctor's Questions and Common Questions support users who are unsure if their symptoms are serious and whether they need medical treatment. Based on Leventhal's Common Sense Model of Self-regulation of health and illness to increase understanding of illness. Treatment Options supports management of distressing symptoms. Based on Bandura's Social Cognitive Theory to increase self-efficacy.*
In what order is it anticipated the components will be used?	*Tunneled components have to be used first. No expectation for menu components.*	*Doctor's Questions first so as to check symptoms.*
What interactive features are available (e.g. forums, videos, printable information)? How long should	*Goal setting component - users with low intended handwashing who select no intention to improve*	*Videos available: Welcome video on home page: 1 minute, 23 seconds, glands: 48 seconds,*

they take to complete?	*receive supportive messages encouraging them to review their choices. An optional print-out page is available.*	*meningitis/septicaemia: 1 minute, 1 second, video sinusitis: 30 seconds. Printable material available but log-data doesn't record whether it was used.*
Are all components/features available to all users throughout the intervention or are some tailored for specific times or users?	*Users self-report amount of handwashing, efficacy and necessity beliefs at the start of sessions 2-4. Users with low handwashing and beliefs are presented with additional tailored information during the tunneled component. Some content is also tailored depending on household type (e.g. for users with children). For dissemination there will be no tailoring by household.*	*Doctor's Questions and Treatment Options are split into cough, sore throat, runny nose & fever.*
Which pages are for collecting self-report measures or for administrative purposes	*Page names starting with 's' or 't' indicate active content.*	*Separate list available for content by individual page.*
Are there specific pages to mark the start and end of sessions?	*First page of tunneled component marks start of each session.*	*Home page is the first page visited.*
Which pages contain BCTs (e.g. information, planning, feedback) and what are they?	*Tunneled pages at the start of each session contain: messages to increase perceived risk, information for explanation of infection transmission by hand, motivation to increase positive attitudes to handwashing, information on viral load and washing hands to reduce infection, goal setting. Tailored pages within tunneled components contain messages to support: habit formation, overcoming barriers, understanding necessity.*	*See components listed previously*
In which sessions are they available?	*Across all 4.*	*N/a*
Can specific BCTs be identified on particular pages or groups of pages? How many groups are there?	*Goal setting in session 1 provides cues and feedback over 4 pages. Tailored pages across sessions 2-4 are grouped by handwashing, necessity, and efficacy. Pages addressing attitudes, barriers and beliefs are used throughout the tailored pages.*	*See components listed previously*
Do any of the pages have response options to collect	*Goal setting pages collect actual and intended*	*Information on symptoms is collected in Doctors*

information in addition to baseline/follow-up measures? What data is collected?	*handwashing plan. However, repeat use within a session is not recorded, only the last entry is stored. For dissemination: amend to capture repeated page viewings and goal entries.*	*Questions.*
2. Accrued data. Data collected during an intervention.		
2.1. Self-report. Users' self-reported responses collected across various stages of the trial.		
When are self-report questionnaires collected?	*Baseline and every 4 weeks up to 12 weeks. For dissemination: create optional follow-up survey completed one week later.*	*Baseline, every 4 weeks up to 24 weeks, during illness and 48 hour follow up.*
What demographic information is available (e.g. age, gender, education)?	*Baseline: Age, gender, qualification, household. For dissemination: add measure for how they heard about the intervention.*	*Baseline: Age, gender, qualification, household, smoker, alcohol, ethnicity.*
Which measures are specifically related to the target behaviour and how often are they collected?	*Baseline, session 2-4, 4 & 12 weeks: Actual and intended handwashing behaviour. Every 4 weeks: User and household illness occurrence. For dissemination: follow-up measures at 1 week for current and intended handwashing behaviour, necessity and efficacy.*	*During illness, every 4 weeks & follow-up (see also external data): occurrence of illness and contact with all NHS.*
Which measures of target behavioral determinants are collected and when?	*Baseline, 4 & 12 weeks: TPB intentions, attitudes, norms, beliefs.*	*Baseline & follow-up: health locus of control, Krantz health opinion survey, TPB. When ill & 48hr: IPQ-R, TPB.*
Are measures of health collected (e.g. conditions which may impact on target behaviour or are co-morbid) and psychosocial factors (e.g. anxiety, illness perception, motivation)?	*Flu vaccination.*	*Baseline, when ill & 48 hour, & follow-up: (Physical) Mobility, self-care, usual activities, pain, anxiety/depression. Baseline (see also external data): health anxiety inventory.*
Are additional measures collected at follow-up (e.g. satisfaction, adherence)?	*Week 8: Satisfaction questions. For dissemination: acceptability e-scale.*	*Follow-up: patient enablement, website satisfaction, problematic experiences of therapy scale.*
2.2. Log-data. Information automatically collected through engagement with an intervention.		
What data is the software platform able to record?	*Time and date, pages viewed & order, time spent on pages, self-report measures.*	*Time and date, pages viewed & order, time spent on pages, self-report measures.*

Are number, date and time of logins available by individual user?	*Yes.*	*Yes.*
Are individuals' total durations of usage accessible?	*Needs to be extracted.*	*Needs to be extracted.*
Are the number and time of usage prompts recorded?	*Overall scheduled timings for emails are available, but not sent times by individual.*	*Overall scheduled timings for emails are available, but not sent times by individual.*
Are there details for which pages were viewed, the sequential order and time spent viewing?	*Yes.*	*Yes.*
2.3. External data. Data collected independently but alongside intervention usage.		
How and where is the data collected?	*Hand collected by the research team from users' GP notes.*	*Hand collected by the research team from users' GP notes.*
What data is collected?	*Visits to a GP for an RTI or gastrointestinal infection during trial period. Antibiotic prescriptions for RTI.*	*Number of GP visits for RTI during trial, for year prior to trial, and co-morbid illnesses. Antibiotic prescriptions for RTI.*
Which of these measures relate to or may impact on the target behaviour?	*Number of GP visits for RTI by user*	*Number of GP visits for RTI during trial, for year prior to trial, and co-morbid illnesses.*
3. Contextual data. Data indirectly related to the running of the intervention which may be influential over usage and analysis.		
3.1. External factors. Structures and events which may influence participation in the intervention.		
How are users recruited to the intervention?	*Recruited via GP using paper consent. For dissemination: move to online consent and capture where users heard about the intervention.*	*Recruited via GP.*
Did any specific large-scale events, with the potential to impact on the intervention, occur during the period of the intervention?	*Outbreak of swine flu prior to trial commencing.*	*N/a*
3.2. Previous theory and findings. Results of behavioral analyses carried out during intervention development (e.g. logic models), and analyses of clinical outcomes if available.		
What are the hypothesized mechanisms of the intervention (e.g. as specified in the intervention's logic model)?	*Handwashing will increase when intentions to wash hands are raised through positive attitudes, norms and beliefs. Increasing perceptions of risk from an RTI will*	*Users with low understanding of their symptoms will be less likely to consult their GP having viewed Doctors Questions and Common Questions. Users with low self-efficacy in their ability to self-manage their illness*

	lead to increased intentions to handwash.	*will be less likely to consult their GP having viewed Treatment Options.* *Level of anxiety may be associated with usage.*
Which factors are identified as important in qualitative research, and can they be related to the variables collected in the trial?	*N/a*	*N/a*
Which variables are identified as relating to outcomes (e.g. behavioral determinants, theoretical constructs, health factors)?	*The largest increase in handwashing occurred at the first session. For dissemination: amend from 4 to 1 session. Perceived risk of infection was a key predictor of intentions to handwash. For dissemination: capture risk at baseline.*	*Analysis of the data from the RCT focused on the outcomes of GP contacts and antibiotic use. No analyses were carried out for usage, behavioral determinants or personal characteristics.*

Appendix C AMUsED framework stage 2 checklist for PRIMIT study and Internet Dr

Selecting usage variables and generating research questions		
Generic questions	Intervention: *Germ Defence*	Intervention: *Internet Dr*
1. Descriptions of usage variables. Which usage variables are relevant to the intervention and in which format (e.g. number of users/sessions, duration, percentage of total, dichotomous)?		
Completing intervention/trial period (stage 1; 1.1 & 2.2).	*Number of users completing tunneled component. Number of pages viewed (including baseline measures, tunneled and individual menu components). Time spent by page.*	*Number of users completing follow-up measures at 24 weeks. Number of users ill during trial period (logged in during illness, self-report at 4 week measures, from GP notes). Total number of pages viewed across the trial and pages with the greatest number of views.* *Total time spent on pages.*
Logins or sessions where the intervention was accessed (stage 1; 1.1 & 2.2).	*N/a*	*Purpose of login: to complete baseline, 4 weekly measures or manage illness. Content viewed, number of pages and time spent when using intervention during illness.*
Date of login and usage.	*N/a*	*n/a*
Time of day of login and usage.	*N/a*	*Proportion of usage outside vs during GP hours.*
Days/weeks of usage (stage 1; 1.1 & 2.2).	*N/a*	*Number of times intervention content was accessed.*
Response to prompts/notifications (stage 1; 1.1 & 2.2).	*Number of users accessing the survey.* *Number of users completing the survey.*	*Number of users who logged in to illness follow-up at 48 hour.*
Features/menu components used (stage 1; 1.1, 1.2 & 2.2).	*Number of users viewing each menu component. Number of pages viewed in menu components. Number of users viewing printout page.*	*Number of users who viewed each of the linked menu components. Component most frequently viewed first. Doctors Questions: Dropout by page. Number of users moving back and forwards through pages. Split of users by symptom. Number of users who saw the advice pages, and type of advice given. Treatment*

		Options & Common Questions: Number of pages viewed. Time spent in this component. Number of users who viewed each of the treatment options. Split of users by symptom. *Number of users viewing each video. Time spend on the video.*
Revisiting components/features (stage 1; 1.1, 1.2 & 2.2).	*Number of users receiving tailored feedback for low or no improvement in behavior who revisit goal setting component. Number of users returning to the start of the tunneled component from the menu page.*	*Number of users re-visiting each component and frequency of revisits. Number of re-visits during illness. Number of users revisiting components with the same/different symptoms. Number of users who login when ill more than once.*
Type of content/BCTs used (excluding administration pages) (stage 1; 1.2 & 2.2).	*User dropout by page (including consent and baseline measures pages).*	*See features/menu components*
Completing ongoing measures (stage 1; 2.1 & 2.2).	*Number of users entering their email address to take part in the survey. Number of questions completed in survey. Number of users completing each question.*	*Number of users completing monthly measures.*
External device usage (stage 1; 2.3).	*N/a*	*N/a*
2. Relationships between usage and participant characteristics. Are user’s demographic, physical or psychosocial characteristics at baseline related to intervention usage?		
Are any characteristics at baseline related to usage?	*Is age, gender or education related to usage? Is type of household related to usage? Are perceived likelihood and/or severity for user or a member of their household related to usage?*	*Is having a co-morbid illness related to usage? Is age associated with different patterns of usage? Is level of anxiety at baseline associated with different patterns of usage?*
Are any contextual factors related to usage (stage 1; 3)?	*Does how users hear about the website relate to usage?*	N/a
Do high/low users differ by other usage factors?	*Is there a relationship between intervention usage and survey usage?*	*Is logging in when ill related to content previously viewed? Do users who self-report GP visits use the intervention differently to users who visited their GP but didn’t report it?*

3. Relationships between usage, behavioral determinants, and target behaviors. Which usage variables are associated with follow-up measures for target behaviour and behavioral determinants? Which usage variables help explain changes in behaviour across the intervention?

Are baseline measures for behavioral determinants/target behavior related to usage? E.g. Is the number of days the intervention is used for related to a behavioral determinant? Do users with low target behavior spend less time on the intervention?	*Is handwashing level at baseline related to usage? Is perceived necessity or efficacy of handwashing related to usage?*	*Is the number of GP visits during year prior to trial related to usage? Are measures for health locus of control, Krantz health opinion survey, or TPB attitudes/norms/intentions at baseline related to usage? Are measures for the IPQ-R, TPB beliefs/intentions at first login during illness related to usage during the period of illness (e.g. viewing content, completing 48hr follow-up).*
Which usage variables are related to behavioral determinants/target behaviors and at follow-up? E.g. Do users who view a group of pages containing a specific BCT score higher/lower for the associated behavioral determinant? Is completing/not completing a particular component associated with target behavior at follow-up? Is the time spent on a session related to target behavior?	*Which usage variables are related to handwashing in the survey? Which usage variables are related to perceived necessity, and efficacy?*	*Is usage related to GP appointments? Is logging in when ill related to GP appointments? Is usage associated with the amount of times a user is ill during the trial? Are measures for health locus of control, Krantz health opinion survey, or TPB attitudes/norms/intentions at follow-up related to usage? Are measures for the IPQ-R, TPB beliefs/intentions at 48hr related to viewing content during illness?*
Is usage associated with measures for acceptability/satisfaction at follow-up? E.g. Are high levels of satisfaction associated with accessing more pages? Do users with low satisfaction spend less time using external devices?	*Which usage variables are associated with scores for user satisfaction, and acceptability e-scale measures?*	*Are scores for patient enablement, website satisfaction, or problematic experiences of therapy scale related to usage?*
Do users who report positive changes in behavioral determinants/target behavior from baseline to follow-up use the intervention differently to those who do not?	*Do users who report increases in handwashing use the website differently to those who don't report increases? Do users who report increases in perceived necessity and efficacy use the intervention differently?*	*Did users with positive changes for measures for health locus of control, Krantz health opinion survey, or TPB attitudes/norms/intentions use the intervention differently? Did users who reported increased positive beliefs at 48hr illness follow-up view different content during illness?*

Are relationships between usage and outcome measures moderated by demographic, psychosocial or health factors?	*Does perceived level of risk moderate the relationship between usage and increased handwashing?*	*Does anxiety moderate relationships between usage and GP visits? Does age moderate the relationship between usage and GP visits?*
What level of usage is necessary for 'effective engagement'?	*Do any of the usage variables have a point of sufficient engagement where maximum behaviour change has occurred?*	*Did users need to see a specific amount or component of content, at a certain time in order to change their behaviour?*

Appendix D AMUsED framework stage 3 checklist for PRIMIT study and Internet Dr

Preparation for analysis		
Generic questions	Intervention: *Germ Defence*	Intervention: *Internet Dr*
1. Resources		
What is the timeframe for completing the analyses?	*By end of 2019.*	*Oct 2018.*
What resources are needed? E.g. additional research time, expertise	*At least 200 users.* *Support with analyses.*	*Support with analyses, confounding variables (e.g. co morbid illnesses), and moderators. Support with visually exploring and extracting data.*
Is a plan of analysis already available? How does the analysis plan developed using the framework compare to that plan? Are changes or updates needed?	*No other plan available.*	*Primary analyses have already been carried out (see stage 1, 3.2), no further plan is available.*
Is ethical clearance in place to carry out usage analyses?	*Yes.*	*Yes.*
2. Selecting types of analysis and analytical software		
Will the usage data be triangulated with qualitative data?	*No.*	*No.*
What analytical tools are available?	*SPSS & LifeGuide Visualisation Tool (LVT).*	*SPSS and LifeGuide Visualisation Tool (LVT).*
Is there sufficient statistical power to answer the planned research questions?	*No, analyses will be exploratory.*	*Analyses using whole intervention group are sufficiently powered, sub-group analyses will not be.*
Can the selected measures of usage be analyzed using the available tools? Is bespoke software necessary (e.g. visualisation techniques)?	*LVT will be necessary for looking at page flow.*	*Sequence of use of components and movement through Doctors Questions will need to be analyzed using LVT.*
3. Data preparation		
When is the data available?	*From spring 2019.*	*Now.*
Is the data raw or has it been used/cleaned previously?	*Raw log-data.*	*Raw log-data.*
How many datasheets are there? Will these need to be amalgamated?	*8 datasheets: 4 log-data sheets for both intervention and survey. Survey data will be matched to intervention data by assigned ID. LVT: Session details will have recorded all necessary follow-up measures. SPSS: Page durations will need to be extracted from page durations sheet and copied*	*5 datasheets: 4 log-data sheets, GP notes. LVT: Session details must contain important self-report measures from user data. Copy primary outcome measures, previous behaviour and co morbid illnesses from GP notes. SPSS: all data for analysis must be available on one data sheet. Relevant measures from GP notes will be copied to session*

	to session data.	*details, along with usage variables from user data and page durations.*
Is the data structured to work with the tools available? What formats are the datasheets in (e.g. excel, .csv) and will they need converting for analysis?	*Datasheets are compatible with SPSS and LVT. Excel sheets will need converting to .csv for LVT, and uploaded to SPSS.*	*Datasheets are compatible with SPSS and LVT. Log data (session details, page flow, page durations, user data) is excel worksheet. These need to be converting to .csv for LVT. GP notes are SPSS.*
What preparation does the data need (e.g. cleaning, anonymizing)?	*Data needs cleaning.*	*Data needs cleaning.*
Are all variables readily available or will they need extracting/transforming/recoding?	*Identified variables will be available. However, totals for time and numbers of pages viewed will need extracting.*	*User data datasheet has a row of data per login so that a user who has logged in 7 times will have 7 rows of data. However, session details datasheet has only 1 row per user. Data must be transformed into 1 row in order to copy across. Totals for time spent and pages viewed will need extracting.*
Is the data in the right format to answer the research questions? Will it need adapting (e.g. continuous variables changed to categorical)?	*Handwashing, necessity, efficacy and perceived risk are continuous scales so will need changing to categorical for high/low group analysis on LVT.*	*Measures of behavioral determinants are measures as continuous variables. For comparisons of usage by high/low groups these will need to be changed to categorical.*

Appendix E AMUSED framework stage 1 checklist for Internet Dr

Familiarisation with the data – identifying variables	Intervention Name: *Internet Dr*
Generic questions by data type	Intervention Details (including dataset)
1. Intervention characteristics. Data for intervention architecture and content.	
1.2. Workflow. Intervention structure and expected participant interaction and navigation through the intervention.	
How many logins/sessions are available?	*Not structured as sessions, access intended during an RTI.*
When are they available?	*All content available at every access. Prompted logins at: baseline, 5 x 4-weekly interim questionnaires, & follow-up questionnaires. Then during illness and 48 hour follow-up.*
Are new sessions released depending on time elapsed or task-completion?	*No.*
Are there limitations on the availability of the intervention?	*No.*
Is the purpose of the session to collect data and/or use the intervention?	*Questions are asked after login to ascertain user's purpose – either to use intervention for illness or to complete interim questionnaires.*
When is the intervention considered to be finished?	*The study finishes after the follow-up questionnaire at 24 weeks. The intervention was finished when users no longer wished to access it.*
What prompts are used to encourage usage (e.g. emails, texts, notifications) and when are they sent?	*Email prompts for 4 weekly data collection. Additional emails are sent after logging in during illness to prompt repeat login after 48 hours.*
Does the intervention contain 'tunnelled' (compulsory) sequences of pages which users have to view in order to move forward?	*'Doctor's Questions' component is tunnelled as symptoms are assessed before providing illness management advice.*
Are users able to select linked sections they wish to view, and avoid others?	*There are 3 linked menu components available from the home page.*
1.2. Content. Content available within the pages of the intervention.	
What are the linked menu sections available?	*'Doctor's Questions'. Split into cough, sore throat, runny nose & fever. Asks about symptoms & provides advice recommending either: self-management, phoning NHS Direct for more advice, or seeking immediate*

	attention via NHS Direct. *'Treatment Options'. Split into cough, sore throat, runny nose & fever. Advice on coping with symptoms: without medication, with medication, and boosting immune system.* *'Common Questions'. 2 sets of FAQs: 'Ask the Internet Dr' - medical questions about illness and treatment, 'Common myths about Colds and Flu' - general questions about illness beliefs.*
What is the aim of each section and are they based on underlying theoretical constructs?	*Doctor's Questions and Common Questions support users who are unsure if their symptoms are serious and whether they need medical treatment. Based on Leventhal's Common Sense Model of Self-regulation of health and illness* [1] *to increase understanding of illness.* *Treatment Options supports management of distressing symptoms. Based on Bandura's Social Cognitive Theory* [2] *to increase self-efficacy.*
In what order is it anticipated the sections will be used?	*No order is necessary but it is anticipated that Doctor's Questions will be viewed first as it is listed first.*
What features are available (e.g. forums, videos, printable information)? How long should they take to complete?	*Videos available: Welcome video on home page: 1 minute, 23 seconds, glands: 48 seconds, meningitis/septicaemia: 1 minute, 1 second, video sinusitis: 30 seconds. Printable material available but log-data doesn't record whether it was used.*
Are all sections/features available to all users throughout the intervention or are some tailored for specific times or users?	*Doctor's Questions and Treatment Options are split into cough, sore throat, runny nose & fever.*
Which pages are for data collection or to support use of the intervention, such as questionnaires or administration of the intervention (e.g. login, password change)?	*See Figure 1 for intervention flow chart showing content by individual page.*
Are there specific pages to mark the start and end of sessions?	*Home page is the first page visited.*
Which pages contain BCTs, such as information, planning, feedback etc...?	*See components listed above.*
What are the underlying BCTs associated with the page?	*N/a*
In which sessions are they available?	*See components listed above.*

Can specific BCTs be identified on particular pages or groups of pages? How many groups are there?	*Information on symptoms is collected in Doctors Questions.*
Do any of the pages have response options to collect additional information? What data is collected?	*See components listed above.*
2. Accrued data. Data collected during an intervention.	
2.1. Self-report. Users' self-reported responses collected across various stages of the trial.	
When are self-report questionnaires collected (e.g. weekly logins, monthly symptom information, follow-up at 6 months)?	*Baseline, every 4 weeks up to 24 weeks, during illness and 48 hour follow up.*
What demographic information is available (e.g. age, gender, education)?	*Baseline: Age, gender, qualification, household, smoker, alcohol, ethnicity.*
Which measures are specifically related to the target behavior and how often are they collected?	*During illness, every 4 weeks & follow-up (see also external data): GP visits & illness occurrence.* *Follow-up: Patient enablement index* [3]
Which measures of beliefs influential on the target behavior are collected and when?	*Baseline & follow-up: health locus of control* [4]*, Krantz health opinion survey* [5]*, TPB* [6]*. When ill & 48hr: IPQ-R* [7]*, TPB* [6]*.*
Are measures collected for health (e.g. conditions which may impact on target behavior or are co-morbid) and psychosocial factors (e.g. anxiety, illness perception, motivation)?	*Baseline, when ill & 48 hour, & follow-up: (Physical) Mobility, self-care, usual activities, pain, anxiety/depression. Baseline (see also external data): health anxiety inventory* [8]*.*
Are additional measures collected at follow-up (e.g. satisfaction, adherence)?	*Follow-up: website satisfaction, problematic experiences of therapy scale* [9]*.*
2.2. Log-data. Information automatically collected through engagement with an intervention.	
What data is the software platform able to record?	*Time and date, pages viewed & order, time spent on pages, self-report measures.*
Are number, date and time of logins available by individual user?	*Yes.*
Are individuals' total durations of usage accessible?	*Need to be extracted.*
Are the number and time of usage prompts recorded?	*Overall scheduled timings for emails are available, but not sent times by individual.*
Are there details for which pages were viewed, the sequential	*Yes.*

order and time spent viewing?	
2.3. External data. Data collected independently but alongside intervention usage.	
How and where is the data collected?	*Hand collected by the research team from users' GP notes.*
What data is collected?	*Number of GP visits for RTI during trial, for year prior to trial, and co-morbid illnesses. Antibiotic prescriptions for RTI.*
Which of these measures relate to or may impact on the target behavior?	*Number of GP visits for RTI during trial, for year prior to trial, and co-morbid illnesses.*
3. Contextual data. Data indirectly related to the running of the intervention which may be influential over usage and analysis.	
3.1. External factors. Structures and events which may influence participation in the intervention.	
How are users recruited to the intervention?	*Recruited via GP.*
Did any specific large-scale events, with the potential to impact on the intervention, occur during the period of the intervention?	*n/a*
3.2. Previous theory and findings. Results of behavioral analyses carried out during intervention development (e.g. logic models), and analyses of clinical outcomes if available.	
What are the hypothesised mechanisms of the intervention (e.g. as specified in the intervention's logic model)?	*Viewing Doctor's Questions and Common Questions will raise users' understanding of their symptoms, leading to them being less likely to consult their GP.* *Viewing Treatment Options will raise users' self-efficacy in their ability to self-manage their illness, leading to them being less likely to consult their GP.* *Level of anxiety may be associated with usage.*
Which factors are identified as important in qualitative research, and can they be related to the variables collected in the trial (e.g. preferences for specific pages)?	*N/a*
Which variables are identified as relating to outcomes (e.g. behavioral determinants, theoretical constructs, health factors)?	*Analysis of the data from the RCT focused on the outcomes of GP contacts and antibiotic use [10]. No analyses were carried out for usage, behavioral determinants or personal characteristics.*

Appendix F AMUSED framework stage 2 checklist for Internet Dr

Selecting usage variables and generating research questions	Intervention Name: Internet Dr
Generic questions	Intervention Specific
1. **Descriptions of usage variables. Which usage variables are relevant to the intervention and in which format (e.g. number of users/sessions, duration, percentage of total, dichotomous)?**	
Completing intervention/trial period (stage1; 1.1 & 2.2). E.g. How many users complete the trial? What is the average time period taken to complete?	*Number of users completing follow-up measures at 24 weeks.* *Number of users ill during trial period (logged in during illness, self-report at 4 week measures, from GP notes).* *Total number of pages viewed across the trial.Total time spent on pages*
Logins or sessions where the intervention was accessed (stage 1; 1.1 & 2.2). E.g. How many users start/complete each login/session? How long does it take to complete each session? How many pages are viewed within the session? Which session has the highest proportion of pages viewed, or duration of time spent on it?	*Purpose of login: to complete baseline, 4 weekly measures or manage illness.* *Content viewed, number of pages and time spent when using intervention during illness.*
Date of login and usage. E.g. When do users login? What time of year? Are there changes in frequency of logins?	*n/a*
Time of day of login and usage. E.g. What time of day is usage? Are users more likely to spend longer on the intervention at certain times?	*Proportion of usage outside vs during GP hours.*
Days/weeks of usage (stage1; 1.1 & 2.2). E.g. For how many days/weeks out of the total is the intervention accessed for? How many times within a week is the intervention accessed? Are there repeated uses within the same day?	*Number of times intervention content was accessed.*
Response to prompts/notifications (e.g. requests to login, email, text, upload data) (stage1; 1.1 & 2.2). E.g. How many responses are sent? How long after receiving notification do users take to log-in or	*Number of users who logged in to illness follow-up at 48hr.*

respond?	
Features/menu sections used (stage 1; 1.1, 1.2 & 2.2). E.g. How many features/sections are accessed? How many users access each feature/section? Which sections/features are completed and by how many users? Which feature/section has the highest proportion of pages viewed? Which feature/section has the highest proportion of time spent on it? What order are the features/sections viewed in? Is this the anticipated order? Which features/sections have the highest proportion of drop-out?	*Number of users who viewed each of the components.* *Component most frequently viewed first.* *Component most frequently viewed first during a login for illness.* *Doctors Questions: Dropout by page. Number of users moving back and forwards through pages. Split of users by symptom. Number of users who saw the advice pages, and type of advice given.* *Treatment Options: Number of pages viewed. Time spent in this section. Number of users who viewed each of the treatment options. Split of users by symptom.* *Common Questions: Number of pages viewed. Time spent in this section. Number of users who viewed the two sections, and each of the questions.* *Number of users viewing each video. Time spend on the video.*
Revisiting sections/features (stage 1; 1.1, 1.2 & 2.2). E.g. Are sections used repeatedly? How many times are they revisited, and for how long? Which sections/features are most revisited?	*Number of users re-visiting each component and frequency of revisits. Number of re-visits during illness.* *Number of users revisiting sections with the same symptoms/different symptoms.* *Number of users who login when ill more than once.*
Type of content/BCTs used (excluding administration pages) (stage 1; 1.2 & 2.2). E.g. How many groups of pages with similar content are accessed and by how many users? How many pages within the group are used? How many users view each page? Which groups of pages have the highest proportion of views? Which pages are viewed at each login, and when is the largest amount of pages viewed? Which pages have higher drop-out?	*See features/menu sections*
Completing ongoing measures (e.g. monthly questionnaires, response options within content pages, uploading information or text	*Number of users logging in when ill, and also viewing content at that time.*

responses) (stage 1; 2.1 & 2.2). E.g. How many users complete ongoing measures? When do they complete them? Do they also access the intervention during that session?	
External device usage (e.g. wearables and other sensor technologies) (stage 1; 2.3). E.g. How much time is spent with the device? How many times is it used? What number of days/weeks is it used for?	*n/a*
2. Relationships between usage and participant characteristics. Are user's demographic, physical or psychosocial characteristics at baseline related to intervention usage?	
Are any characteristics at baseline related to usage? E.g. Is education level associated with number of pages viewed? Is anxiety associated with revisiting features? Is current health related to usage of external devices? Are users who spend more time on the intervention older than those who spend less time? Which characteristics are associated with drop-out?	*Are age or gender associated with different patterns of usage?* *Is level of anxiety at baseline associated with different patterns of usage?* *Is having visited a GP in the past year associated with usage?*
Are any contextual factors related to usage (stage1; 3)? E.g. Does manner of recruitment relate to usage?	n/a
Do high/low users differ by other usage factors? E.g. Do users with higher levels of usage view more types of content than users with lower levels? Do users who complete more sessions also access a specific feature more often? Is usage of an external device related to intervention usage?	*Is logging in when ill related to content previously viewed?* *Do users who self-report GP visits use the intervention differently to users who visited their GP but didn't report it?*
3. Relationships between usage, target behaviours and behavioural determinants. Which usage variables are associated with follow-up measures for target behaviour and behavioural determinants? Which usage variables help explain changes in behaviour across the intervention?	
Are baseline measures for target behaviour and behavioural influencers related to usage? E.g. Do users with low target behaviour spend less time on the intervention? Do users with high target behaviour revisit a specific feature? Is the number of days the intervention is used for	*Are measures of believed necessity for HCP at baseline associated with usage?* *Are measures of believed necessity for HCP associated with usage during the period of illness (e.g. viewing content, completing 48hr follow-up).*

related to a behavioural determinant?	
Which usage variables are related to target behaviours and behavioural determinants at follow-up? E.g. Is completing/not completing a particular section associated with target behaviour at follow-up? Is the time spent on a session related to target behaviour? Do users with high target behaviour at follow-up complete a specific section compared to users with low target behaviour? Do users who view a group of pages containing a specific BCT score higher/lower for the associated behavioural determinant?	*Is usage related to enablement? Is logging in when ill related to enablement scores?* *Is usage associated with the amount of times a user is ill during the trial?* *Are changes in measures of believed necessity for HCP across the study associated with usage?*
Is usage associated with measures for satisfaction at follow-up? E.g. Are high levels of satisfaction associated with accessing more pages? Do users with low satisfaction spend less time using external devices?	*Are scores for website satisfaction or usage experience associated with usage?*
Do users who report positive changes in target behaviour/behavioural determinants from baseline to follow-up use the intervention differently to those who do not? E.g. Do users who report positive behaviour change spend more time on the intervention? Do users who report increases in a behavioural determinant view more pages from a specific section containing a relevant BCT?	*Did users with lowered beliefs in necessity of HCP use the intervention differently?* *Did users who reported increased positive beliefs at 48hr illness follow-up view different content during illness?*
Are relationships between usage and outcome measures moderated by demographic, psychosocial or health factors? E.g. Is the relationship between time spent on the intervention and behaviour altered when moderated by anxiety?	*Does anxiety moderate relationships between usage and enablement?* *Does age moderate the relationship between usage and enablement?*
What level of usage is necessary for ‘effective engagement’? E.g. Do outcome measures plateau after viewing certain content, or after a certain amount of time or sessions completed?	*Did users need to see a specific amount or section of content, at a certain time in order to change their behaviour?*

Appendix G AMUSED framework stage 3 checklist for Internet Dr

Preparation for analysis	Intervention Name:
Generic questions	Intervention Details
1. Resources	
What is the timeframe for completing the analyses?	*August 2019*
What resources are needed? E.g. additional research time, expertise.	*Expertise on types of analyses, confounding variables (e.g. co morbid illnesses), and moderators. Support with visualisation tool, and help extracting data.*
Is a plan of analysis already available? How does the analysis plan developed using the framework compare to that plan? Are changes or updates needed?	*Primary analyses have already been carried out (see stage 1, 3.2), no further plan is available.*
2. Selecting types of analysis and analytical software	
Will the usage data be triangulated with qualitative data?	*No*
Is there sufficient statistical power to answer the planned research questions?	*Analyses using whole intervention group are sufficiently powered, sub-group analyses may not be.*
What analytical tools are available?	*SPSS and LifeGuide Visualisation Tool (LVT)*
Can the selected measures of usage be analysed using the available tools? Is bespoke software necessary (e.g. visualisation techniques)?	*Sequence of use of sections and movement through Doctors Questions will need to be analysed using LVT*
3. Data preparation	
When is the data available?	*Now*
How many datasheets are there? Will these need to be amalgamated?	*7 datasheets: Log data (session details, page flow, page durations, user data), GP notes, two SPSS datasheets from previous analysis.* *LVT: Need individual sheets for LVT. Session details must contain*

	important self-report measures from user data. Most measures have already been saved, any measures outstanding will need to be copied across. Need primary outcome measures, previous behaviour and co morbid illnesses from GP notes. *SPSS: all data for analysis must be available on one data sheet. The two previous SPSS sheets contain some but not all measures relevant to usage. Relevant measures from GP notes will be copied to session details, along with usage variables from user data and page durations.*
What format is the datasheet in (e.g. excel, .csv)? Will it need converting for analysis?	*Log data (session details, page flow, page durations, user data) is excel. These will need to be converted from worksheets to .csv for LVT.* *GP notes SPSS.* *SPSS datasheets from previous analysis.*
Is the data structured to work with the tools available? What changes or data cleaning are needed?	*Data has already been collected and is compatible with LVT.*
What preparation does the data need (e.g. cleaning, anonymising)?	*Data needs cleaning and anonymising.*
Are all variables readily available or will they need extracting/transforming/recoding?	*In the user data datasheet each login is recorded on a new row so that a user who has logged in 7 times will have 7 rows of data. However, in the session details datasheet each user has only 1 row of data. To make it possible to copy variables across from user data to session details, they must first be transformed into 1 row of data. Code will be needed to extract data for total intervention time and pages used, and temporal analyses.*
Is the data in the right format to answer the research questions? Will it need adapting (e.g. continuous variables changed to categorical)?	*Behavioural constructs are measured as continuous variables. For comparisons of usage by high/low groups these will need to be changed to categorical.*

Appendix H Internet Dr study: psychological measures collected at baseline and follow-up

Collected online across trial period			
Measure	***Description***	***Baseline***	***24 weeks***
Health Anxiety Inventory[21]	14 items, 4 response options numbered 0-3, where 3 is highest level of anxiety (e.g. I do not worry about my health/I spend most of my time worrying about my health).	✓	
Health Locus of Control, powerful others B scale[22]	3 items, 0-7 Likert scale ranging from 'strongly agree' to 'strongly disagree' where low scores indicate high external locus of control/greater reliance on HCP (e.g. If I see an excellent doctor regularly, I am less likely to have health problems).	✓	✓
Krantz Health Opinion Survey, behavioral involvement scale[23]	7 items. 0-7 Likert scale ranging from 'strongly agree' to 'strongly disagree'. Reverse scored where necessary so that low scores indicate stronger reliance upon HCP (e.g. Learning how to cure some of your own illness without contacting a physician may create more harm than good).	✓	✓
Theory of Planned Behaviour (TPB)[24]	*Attitudes*. 2 items. 2 response options (e.g. Getting and following advice from this website will be/was: useful/useless).	✓	✓
	Subjective Norms. 2 items, 0-7 Likert scale, ranging from 'strongly disagree' to 'strongly agree', where high scores indicate positive norms to use and follow the website advice (e.g. My family and friends think that it is/was sensible for me to get and follow advice from this website).	✓	✓
	Perceived Behavioural Control (PBC). 2 items, 0-7 Likert scale, ranging from 'strongly disagree' to 'strongly agree', where high scores indicate positive beliefs to use and follow the website advice (e.g. It will be/was easy for me to get/follow advice from this website).	✓	✓
	Intentions. 2 items, 0-7 Likert scale, ranging from 'strongly disagree' to 'strongly agree', where high scores indicate higher intentions to use and follow the advice (e.g. I intend to get and follow advice from	✓	

	this website).	
Problematic Experiences of Therapy Scale (PETS)[25]	*Symptoms*. 3 items, 1-5 Likert scale ranging from ‘agree strongly’ to ‘disagree strongly’, where low scores indicate a greater degree of problem (e.g. I did not use/follow the Internet Doctor advice because it made my symptoms worse).	✓
	Uncertainty about how to use the intervention. 3 items, 1-5 Likert scale ranging from ‘agree strongly’ to ‘disagree strongly’, where low scores indicate a greater degree of problem (e.g. I could not use/follow the Internet Doctor advice because I was unsure how to do it properly).	✓
	Doubts about intervention efficacy. 3 items, 1-5 Likert scale ranging from ‘agree strongly’ to ‘disagree strongly’, where low scores indicate a greater degree of problem (e.g. I did not use/follow the Internet Doctor advice because I was not sure if it was helping).	✓
	Practical problems. 5 items, 1-5 Likert scale ranging from ‘agree strongly’ to ‘disagree strongly’, where low scores indicate a greater degree of problem (e.g. I found it difficult to remember to use/follow the Internet Doctor advice).	✓
Patient Enablement Index (PEI)[8]	6 items. Thinking about the kinds of symptoms we have asked about in this study (e.g. cough, sore throat, sinus pain, runny nose, cold and flu symptoms), compared with before you took part in this study, do you feel you are: Able to cope with life when you have these symptoms? Able to understand these symptoms? Able to cope with these symptoms? Able to keep yourself healthy? Confident about your health when you have these symptoms? Able to help yourself when you have these symptoms? 0-2 Likert scale, where 0 = ‘same or less’, 1 = ‘better’, 2 = ‘much better’. High scores indicate better enablement.	✓

Appendix I PRIMIT/Germ Defence: Table of results for logistic regression analysis of baseline user characteristics predicting completion of core section of intervention

Predictors	***β***	***SE***	**Wald**	***df***	***P***	**Exp(B)**	**95% CI**	
							Lower	**Upper**
Gender								
PRIMIT Study	-.079	.083	.840	1	.359	.927	.788	1.090
Germ Defence	.445	.362	1.510	1	.219	1.561	.767	3.175
Age								
PRIMIT Study	.000	.003	.005	1	.942	1.000	.994	1.006
Germ Defence	.009	.013	.458	1	.449	1.009	.983	1.035
Current daily handwashing								
PRIMIT Study	.111	.072	2.386	1	.122	1.117	.971	1.286
Germ Defence	.228	.225	1.029	1	.310	1.256	.808	1.953
Intended daily handwashing								
PRIMIT Study	-.068	.075	.823	1	.364	.934	.806	1.082
Germ Defence	-.002	.229	.000	1	.994	.998	.638	1.563
Perceived likelihood of user becoming ill								
PRIMIT Study	.006	.024	.060	1	.806	1.006	.959	1.055
Germ Defence	.214	.115	3.431	1	.064	1.238	.988	1.553
Perceived severity for user								
PRIMIT Study	-	-	-	-	-	-	-	-
Germ Defence	-.131	.107	1.511	1	.219	.877	.711	1.081

SE = standard error, *df* = degrees of freedom, *p* = significance, OR = odds ratio, CI = confidence interval

Appendix J PRIMIT/Germ Defence: Mean differences and 95% confidence intervals for goal-setting handwashing frequencies

Table 1. Comparison of mean scores for current vs intended levels of handwashing within the same study/intervention

Situation	**PRIMIT study**			**Germ Defence**		
	MD	**95% CI**		***MD***	**95% CI**	
		Lower	**Upper**		**Lower**	**Upper**
C1 (before eating meals)	**.654**	**.634**	**.673**	**.718**	**.557**	**.878**
C2 (before eating snacks)	**1.040**	**1.016**	**1.063**	**.949**	**.766**	**1.132**
C3 (after going to the toilet)	.079	.072	.086	.040	-.003	.082
C4 (coming in to the house)	**.930**	**.907**	**.953**	**.661**	**.502**	**.820**
C5 (after being close to someone who is ill)	**.902**	**.878**	**.926**	**.768**	**.575**	**.961**
C6 (after sneezing or coughing)	**.995**	**.972**	**1.019**	**.842**	**.675**	**1.008**
C7 (after touching something with germs on)	-	-	-	.092	.192	.292

Table 2. Comparison of mean scores between PRIMIT vs Germ Defence studies for current behaviour and intended behaviour

Situation	**Current**			**Intended**		
	MD	**95% CI**		***MD***	**95% CI**	
		Lower	**Upper**		**Lower**	**Upper**
C1 (before eating meals)	**.305**	**.108**	**.502**	.261	.112	.409
C2 (before eating snacks)	.146	-.062	.355	.237	.047	.428
C3 (after going to the toilet)	-.063	-.148	.022	-.021	-.084	.041
C4 (coming in to the house)	-.229	-.449	-.009	.041	-.149	.231
C5 (after being close to someone who is ill)	.084	-.140	.308	.207	.035	.380
C6 (after sneezing or coughing)	.242	0.32	.453	**.381**	**.206**	**.557**
C7 (after touching something with germs on)	-	-	-	-	-	-

Table 3. Comparison of mean scores between HCPs vs non-HCPs for Germ Defence

Situation	**Current**			**Intended**		
	MD	**95% CI**		***MD***	**95% CI**	
		Lower	**Upper**		**Lower**	**Upper**
C1 (before eating meals)	.044	-.422	.510	-.096	-.479	.288
C2 (before eating snacks)	.117	-.336	.570	.204	-.247	.655
C3 (after going to the toilet)	-.109	-.258	.039	-.052	-.161	.056
C4 (coming in to the house)	-.115	-.607	.377	-.203	-.633	.227
C5 (after being close to someone who is ill)	.216	-.355	.788	.150	-.321	.620
C6 (after sneezing or coughing)	.149	-.319	.617	.055	-.403	.512
C7 (after touching something with germs on)	.052	-.254	.358	.009	-.189	.207

Practically significant results (*MD*>0.3) are in bold.

List of References

Ainsworth, B., Miller, S., Denison-Day, J., Stuart, B., Groot, J., Rice, C., Bostock, J., Hu, X., Morton K., Towler, L., Moore, M., Willcox, M., Chadborn, T., Gold, N., Amlôt, R., Little, P., & Yardley, L. (2020). Infection control behaviour at home during the COVID-19 Pandemic: Observational study of a web-based behavioural Intervention (Germ Defence). *Journal of Medical Internet Research, 23*(2), e22197. https://doi.org/10.2196/22197

Ainsworth, B., Steele, M., Stuart, B., Joseph, J., Miller, S., Morrison, L., Little, P., & Yardley, L. (2017). Using an analysis of behavior change to inform effective digital intervention design: How did the PRIMIT website change hand hygiene behavior bcross 8993 users? *Annals of Behavioral Medicin*e, *51*(3), 423-431. https://doi.org/10.1007/s12160-016-9866-9

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Ajzen, I. (2002). Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology*, *32*(4), 665–683. https://doi.org/10.1111/j.1559-1816.2002.tb00236.x

Alshurafa, N., Jain, J., Alharbi, R., Iakovlev, G., Spring, B., & Pfammatter, A. (2018). Is More Always Better? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(4), 1–26. https://doi.org/10.1145/3287031

Arden-Close, E. J., Smith, E., Bradbury, K., Morrison, L., Dennison, L., Michaelides, D., & Yardley, L. (2015). A visualization tool to analyse usage of web-based interventions: The example of Positive Online Weight Reduction (POWeR). *JMIR Human Factors*, *2*(1), e8. https://doi.org/10.2196/humanfactors.4310

Arigo, D., Jake-Schoffman, D. E., Wolin, K., Beckjord, E., Hekler, E. B., & Pagoto, S. L. (2019). The history and future of digital health in the field of behavioral medicine. *Journal of Behavioral Medicine*, *42*(1), 67-83. https://doi.or/10.1007/s10865-018-9966-z.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Band, R., Bradbury, K., Morton, K., May, C., Michie, S., Mair, F. S., Murray, E., McManus, R., Little, R., & Yardley, L. (2017). Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. *Implementation Science*, *12*(25). https://doi.org/10.1186/s13012-017-0553-4

Bandura, A (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Eds.), *Self-efficacy beliefs of adolescents*. Information Age publishing. ISBN:1593113667

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W H Freeman/Times Books/Henry Holt & Co. ISBN: 9780716728504.

Banks, I. (2010). Self care of minor ailments: A survey of consumer and healthcare professional beliefs and behaviour*. Self Care, 1,* 1-13. https://selfcarejournal.com/wp-content/uploads/2015/09/Banks-1.1.1-13pdf.pdf

Beard, E., West, R., Michie, S., & Brown, J. (2016). Association between electronic cigarette use and changes in quit attempts, success of quit attempts, use of smoking cessation pharmacotherapy, and use of stop smoking services in England: time series analysis of population trends. *BMJ*, *354*, i4645. https://doi.org/10.1136/bmj.i4645

Beintner, I., Vollert, B., Zarski, A.-C., Bolinski, F., Musiat, P., Görlich, D., Ebert, D. D., & Jacobi, C. (2019). Adherence Reporting in Randomized Controlled Trials Examining Manualized Multisession Online Interventions: Systematic Review of Practices and Proposal for Reporting Standards. *Journal of Medical Internet Research*, *21*(8), e14181. https://doi.org/10.2196/14181

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*, *75*(12), 2299–2306. https://doi.org/10.1016/J.SOCSCIMED.2012.08.032

Bricker, J. B., Sridharan, V., Zhu, Y., Mull, K. E., Heffner, J. L., Watson, N. L., McClure, J. B., & Di, C. (2018). Trajectories of 12-month usage patterns for two smoking cessation websites: Exploring how users engage over time. *Journal of Medical Internet Research*, *20*(4). https://doi.org/10.2196/10143

Centers for Disease Control and Prevention (2020). *Handwashing: Clean Hands Save Lives*. https://www.cdc.gov/handwashing/index.html

Cheng, K. K. F., & Metcalfe, A. (2018). Qualitative Methods and Process Evaluation in Clinical Trials Context. *International Journal of Qualitative Methods*, *17*(1), 160940691877421. https://doi.org/10.1177/1609406918774212

Clay, R. A. (2010). More than one way to measure. *American Psychological Association*, *41*(8), 52. https://www.apa.org/monitor/2010/09/trials

Coa, K. I., Wiseman, K. P., Higgins, B., & Augustson, E. (2019). Associations between engagement and outcomes in the SmokefreeTXT program: A growth mixture modeling analysis. *Nicotine & Tobacco Research*, *21*(5), 663–669. https://doi.org/10.1093/ntr/nty073

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American Journal of Preventive Medicine*, *32*(5 Suppl), S112-8. https://doi.org/10.1016/j.amepre.2007.01.022

Colquhoun, H. L., Levac, D., O, K. K., Straus, S., Tricco, A. C., Perrier, L., Kastner, M., & Moher, D. (2014). Scoping reviews: Time for clarity in definition, methods and reporting. *Journal of Clinical Epidemiology*, (12), 1291–1294. https://doi.org/10.1016/j.jclinepi.2014.03.013

Couper, M. P., Alexander, G. L., Zhang, N., Little, R. J., Maddy, N., Nowak, M. A.,McClure, J. B., Calvi, J. J., Rolnick, S. J., Stopponi, M. A., & Cole Johnson, C. (2010). Engagement and Retention: Measuring Breadth and Depth of Participant Use of an Online Intervention. *Journal of Medical Internet Research*, *12*(4), e52. https://doi.org/10.2196/jmir.1430

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M., Medical Research Council Guidance. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, *29*(337), a1655. https://doi.org/10.1136/BMJ.A1655

Danaher, B. G., Boles, S. M., Akers, L., Gordon, J. S., & Severson, H. H. (2006). Defining participant exposure measures in Web-based health behavior change programs. *Journal of Medical Internet Research*, *8*(3), e15. https://doi.org/10.2196/jmir.8.3.e15

Department of Health and Social Care (2020). *Public information campaign focuses on handwashing*. UK Government. https://www.gov.uk/government/news/public-information-campaign-focuses-on-handwashing.

Department of Health and Social Care (2011). *Catch it, Bin it, Kill it - Respiratory and hand hygiene campaign (2011)*. UK Government. https://www.gov.uk/government/publications/catch-it-bin-it-kill-it-respiratory-and-hand-hygiene-campaign-2011.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, *210*, 2–21.

https://doi.org/10.1016/j.socscimed.2017.12.005

Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies. *Journal of Medical Internet Research*, *13*(3), e52. https://doi.org/10.2196/jmir.1772

Donkin, L., Hickie, I. B., Christensen, H., Naismith, S. L., Neal, B., Cockayne, N. L., & Glozier, N. (2013). Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: Randomized controlled trial. *Journal of Medical Internet Research*, *15*(10). https://doi.org/10.2196/jmir.2771

Dunphy, E., Button, K., Hamilton, F., Williams, J., Spasic, I., & Murray, E. (2021). Feasibility randomised controlled trial comparing TRAK-ACL digital rehabilitation intervention plus treatment as usual versus treatment as usual for patients following anterior cruciate ligament reconstruction. *BMJ Open Sport & Exercise Medicine*, *7*(2), e001002. https://doi.org/10.1136/bmjsem-2020-001002

Elbert, N. J., van Os-Medendorp, H., van Renselaar, W., Ekeland, A. G., Hakkaart-van Roijen, L., Raat, H., Nijsten, T. E. C., & Pasmans, S. G. M. A. (2014). Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. *Journal of Medical Internet Research*, *16*(4), e110. https://doi.org/10.2196/jmir.2790

Enrique, A., Palacios, J. E., Ryan, H., & Richards, D. (2019). Exploring the Relationship Between Usage and Outcomes of an Internet-Based Intervention for Individuals With Depressive Symptoms: Secondary Analysis of Data From a Randomized Controlled Trial. *Journal of Medical Internet Research*, *21*(8), e12775. https://doi.org/10.2196/12775

Everitt, H. A., Landau, S., O'Reilly, G., Sibelli, A., Hughes, S., Windgassen, S., Holland, R., Little, P., McCrone, P., Bishop, F., Goldsmith, K., Coleman, N., Logan, R., Chalder, T., Moss-Morris, R., & ACTIB trial group (2019). Assessing telephone-delivered cognitive-behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial. *Gut*, *68*, 1613–1623. https://doi.org/10.1136/gutjnl-2018-317805

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*(1). https://doi.org/10.2196/jmir.7.1.e11

Eysenbach, G., & CONSORT-EHEALTH Group (2011). CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal* *of Medical Internet Research*, *13*(4), e126. https://doi.org/10.2196/jmir.1923

Fleming, T., Bavin, L., Lucassen, M., Stasiak, K., Hopkins, S., & Merry, S. (2018). Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *Journal of Medical Internet Research*, *20*(6). https://doi.org/10.2196/jmir.9275

Fletcher, A., Jamal, F., Moore, G., Evans, R. E., Murphy, S., & Bonell, C. (2016). Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. *Evaluation*, *22*(3), 286–303. https://doi.org/10.1177/1356389016652743

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89*(9), 1322-1327. https://doi.org/10.2105/AJPH.89.9.1322

Gold, N., Of Oxford, U., Hu, X.-Y., Amlôt, R., Chadborn, T., & Yardley, L. (2020). Effectiveness of digital interventions to improve household and community infection prevention and control behaviours and to reduce incidence of respiratory and/or gastro-intestinal infections: A rapid systematic review. *BMC Public Health, 21*(1180). https://doi.org/10.1186/s12889-021-11150-8

Leventhal, H., Brissette, I., Leventhal, E. A. (2003). The common-sense model of self-regulation of health and illness. In L. D. Cameron, & H. A. Leventhal (Eds.), *The self-regualtion of health and illness behavoiur* (pp. 42-65). Routledge. ISBN: 9780415297011

Hamine, S., Gerth-Guyette, E., Faulx, D., Green, B. B., & Ginsburg, A. S. (2015). Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.3951

Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity, 16*(31). https://doi.org/10.1186/s12966-019-0792-7

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, B., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A., & Michie, S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*(g1687).

https://doi.org/10.1136/bmj.g1687

Horwood, J., Chalder, M., Ainsworth, B., Denison-Day, J., de Vocht, F., Elwenspoek, M. M. C., Craggs, P., Denholm, R., Sterne, J., Rice, C., Miller, S., Stuart, B., Little, P., Moore, M., Wilcox, M., Macleod, J., Gullford, M., Morton, K., Towler, L., … Yardley, L. (2021). Primary Care implementation of Germ Defence, a digital behaviour change intervention to improve household infection control during the COVID-19 pandemic: A structured summary of a study protocol for a randomised controlled trial. *Trials*, *22*(1), 263. https://doi.org/10.1186/s13063-021-05188-7

Howie, J. G., Heaney, D. J., Maxwell, M., & Walker, J. J. (1998). A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. *Family Practice*, *15*(2), 165–171. https://doi.org/10.1093/fampra/15.2.165

Issom, D. Z., Henriksen, A., Woldaregay, A. Z., Rochat, J., Lovis, C., & Hartvigsen, G. (2020). Factors influencing motivation and engagement in mobile health among patients with sickle cell disease in low-prevalence, high-income countries: Qualitative exploration of patient requirements. *JMIR Human Factors*, *7*(1). https://doi.org/10.2196/14599

Kelders, S., Kip, H., & Greeff, J. (2020). Psychometric evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation Study. *Journal of Medical Internet Research, 22*(10), e17757. https://doi.org/10.2196/17757

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Van Gemert-Pijnen, J. E. (2012). Persuasive System Design Does Matter: a Systematic Review of Adherence to Web-based Interventions. *Journal of Medical Internet Research*, *14*(6), e152. https://doi.org/10.2196/jmir.2104

Kirby, S., Donovan-Hall, M., & Yardley, L. (2014). Measuring barriers to adherence: validation of the problematic experiences of therapy scale. *Disability and Rehabilitation*, *36*(22), 1924–1929. https://doi.org/10.3109/09638288.2013.876106

Koneska, E., Appelbe, D., Williamson, P. R., & Dodd, S. (2020). Usage metrics of web-based interventions evaluated in randomized controlled trials: Systematic review. *Journal of Medical Internet Research*, *22*(4). https://doi.org/10.2196/15474

Krantz, D. S., Baum, A., Wideman, M. V. (1980). Assessment of Preferences for Self-Treatment and Information in Health Care. *Journal of Personality and Social Psychology*, *39*(5), 977–990.

Lau, J. T. F., Kim, J. H., Tsui, H., & Griffiths, S. (2007). Anticipated and current preventive behaviors in response to an anticipated human-to-human H5N1 epidemic in the Hong Kong Chinese general population. *BMC Infectious Diseases*, *7*(1), 18. https://doi.org/10.1186/1471-2334-7-18

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation Science*, *5*(1), 69. https://doi.org/10.1186/1748-5908-5-69

Little, P., Stuart, B., Hobbs, F. D. R., Moore, M., Barnett, J., Popoola, D., Middleton, K., Kelly, J., Mullee, M., Raftery, J., Yao, G., Carmen, W., Fleming, D., Stokes-Lampard, H., Williamson, I., Joseph, J., Miller, S., & Yardley, L. (2015). An internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): A primary care randomised trial. *The Lancet*, *386*(10004) 1631-1639. https://doi.org/10.1016/S0140-6736(15)60127-1

Little, Paul, Stuart, B., Andreou, P., McDermott, L., Joseph, J., Mullee, M., Moore, M., Broomfield, S., Thomas, T., & Yardley, L. (2016). Primary care randomised controlled trial of a tailored interactive website for the self-management of respiratory infections (Internet Doctor). *BMJ Open*, *6*(4), e009769. https://doi.org/10.1136/bmjopen-2015-009769

Michie, S., Carey, R. N., Johnston, M., Rothman, A. J., de Bruin, M., Kelly, M. P., & Connell, L. E. (2016). From Theory-Inspired to Theory-Based Interventions: A Protocol for Developing and Testing a Methodology for Linking Behaviour Change Techniques to Theoretical Mechanisms of Action. *Annals of Behavioral Medicine*, 1–12. https://doi.org/10.1007/s12160-016-9816-6

Michie, S., Johnston, M., Francis, J., Hardeman, W., & Eccles, M. (2008). From Theory to Intervention: Mapping Theoretically Derived Behavioural Determinants to Behaviour Change Techniques. *Applied Psychology*, *57*(4), 660–680. https://doi.org/10.1111/j.1464-0597.2008.00341.x

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, James., & Wood, C. E. (2013). 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*, *46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Michie, S., Yardley, L., West, R., Patrick, K., & Greaves, F. (2017). Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. *Journal of Medical Internet Research*, *19*(6), e232. https://doi.org/10.2196/jmir.7126

Miller, S., Ainsworth, B., Yardley, L., Milton, A., Weal, M., Smith, P., & Morrison, L. (2019). A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. *Journal of Medical Internet Research*, *21*(2), e10966. https://doi.org/10.2196/10966

Miller, S., Yardley, L., & Little, P. (2012). Development of an intervention to reduce transmission of respiratory infections and pandemic flu: Measuring and predicting hand-washing intentions. *Psychology, Health & Medicine, 17*(1), 59-81. https://doi.org/10.1080/13548506.2011.564188

Mohr, D. C., Schueller, S. M., Montague, E., Burns, M. N., & Rashidi, P. (2014). The behavioral intervention technology model: An integrated conceptual and technological framework for ehealth and mhealth interventions. *Journal of Medical Internet Research 16*(6), e146. https://doi.org/10.2196/jmir.3077

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Cooper, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2014). Process evaluation in complex public health intervention studies: the need for guidance. *Journal of Epidemiology and Community Health*, *68*(2), 101–102. https://doi.org/10.1136/jech-2013-202869

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*, h1258. https://doi.org/10.1136/BMJ.H1258

Morrison, C., & Doherty, G. (2014). Analyzing engagement in a web-based intervention platform through visualizing log-data. *Journal of Medical Internet Research*, *16*(11), e252. https://doi.org/10.2196/jmir.3575

Morrison, L. G., Geraghty, A. W. A., Lloyd, S., Goodman, N., Michaelides, D. T., Hargood, C., Weal, M., & Yardley, L. (2018). Comparing usage of a web and app stress management intervention: An observational study. *Internet Interventions*, *12*, 74–82. https://doi.org/10.1016/J.INVENT.2018.03.006

Morrison, L. G., & Yardley, L. (2009). What infection control measures will people carry out to reduce transmission of pandemic influenza? A focus group study. *BMC Public Health*, *9*(1), 258. https://doi.org/10.1186/1471-2458-9-258

Morton, K., Ainsworth, B., Miller, S., Rice, C., Bostock, J., Denison-Day, J., Towler, L., Groot, J., Moore, M., Willcox, M., Chadborn, T., Amlot, R., Gold, N., Little, P., & Yardley, L. (2021). Adapting Behavioral Interventions for a Changing Public Health Context: A Worked Example of Implementing a Digital Intervention During a Global Pandemic Using Rapid Optimisation Methods. *Frontiers in Public Health*, *9*, 668197. https://doi.org/10.3389/fpubh.2021.668197

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Murray, E., Treweek, S., Pope, C., MacFarlane, A., Ballini, L., Dowrick, C., Finch, T., Kennedy, A., Mair, Frances., O'Donnell, C., Ong, B. N., Rapley, T., Rogers, A., & May, C. (2010). Normalisation process theory: A framework for developing, evaluating and implementing complex interventions. *BMC Medicine*, *8*, 63. https://doi.org/10.1186/1741-7015-8-63

Nazareth, I., & Murray, E. (2010). Promoting self care for minor illness. *BMJ*, *340*, c2913. https://doi.org/10.1136/BMJ.C2913

Nelson, L. A., Coston, T. D., Cherrington, A. L., & Osborn, C. Y. (2016). Patterns of User Engagement with Mobile- and Web-Delivered Self-Care Interventions for Adults with T2DM: A Review of the Literature. *Current Diabetes Reports 16*(7). https://doi.org/10.1007/s11892-016-0755-1

Nelson, L. A., Spieker, A., Greevy, R., LeStourgeon, L. M., Wallston, K. A., & Mayberry, L. S. (2020). User engagement among diverse adults in a 12-month text message–delivered diabetes support intervention: Results from a randomized controlled trial. *JMIR MHealth and UHealth*, *8*(7). https://doi.org/10.2196/17534

Newman, M. G., Szkodny, L. E., Llera, S. J., & Przeworski, A. (2011). A review of technology-assisted self-help and minimal contact therapies for drug and alcohol abuse and smoking addiction: Is human contact necessary for therapeutic efficacy? *Clinical Psychology Review*, *31*(1), 178–186. https://doi.org/10.1016/J.CPR.2010.10.002

Nurmi, J., Knittle, K., Ginchev, T., Khattak, F., Helf, C., Zwickl, P., Castellano-Tejedor, C., Lusilla-Palacios, P., Costa-Requena, J., Ravaja, N., & Haukkala, A. (2020). Engaging users in the behavior change process with digitalized motivational interviewing and gamification: Development and feasibility testing of the precious app. *JMIR MHealth and UHealth*, *8*(1), e12884. https://doi.org/10.2196/12884

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. https://doi.org/10.1002/asi.20801

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*(1), 54–74. https://doi.org/10.1037/0033-2909.124.1.54

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Paz Castro, R., Haug, S., Filler, A., Kowatsch, T., & Schaub, M. P. (2017). Engagement Within a Mobile Phone–Based Smoking Cessation Intervention for Adolescents and its Association With Participant Characteristics and Outcomes. *Journal of Medical Internet Research*, *19*(11), e356. https://doi.org/10.2196/jmir.7928

Perski, O., Blandford, A., Garnett, C., Crane, D., West, R., Michie, S. (2020). A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". *Translational Behavioural Medicine, 10*(1), 267-277. https://doi.org/10.1093/tbm/ibz039

Perski, O, Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine, 7*(2), 254-267. https://doi.org/10.1007/s13142-016-0453-1

Perski, O, Lumsden, J., Garnett, C., Blandford, A., West, R., & Michie, S. (2019). Assessing the psychometric properties of the digital behavior change intervention engagement scale in users of an app for reducing alcohol consumption: Evaluation study. *Journal of Medical Internet Research*, *21*(11), e16197. https://doi.org/10.2196/16197

Perski, O, Naughton, F., Garnett, C., Blandford, A., Beard, E., West, R., & Michie, S. (2019). Do Daily Fluctuations in Psychological and App-Related Variables Predict Engagement With an Alcohol Reduction App? A Series of N-Of-1 Studies. *JMIR MHealth and UHealth*, *7*(10), e14098. https://doi.org/10.2196/14098

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015).

Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare*, *13*(3), 141–146. https://doi.org/10.1097/XEB.0000000000000050

Pham, Q., Graham, G., Carrion, C., Morita, P. P., Seto, E., Stinson, J. N., & Cafazzo, J. A. (2019). A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: Scoping review. *JMIR MHealth and UHealth*, *7*(1). https://doi.org/10.2196/11941

Pham, Q., Shaw, J., Morita, P. P., Seto, E., Stinson, J. N., & Cafazzo, J. A. (2019). The Service of Research Analytics to Optimize Digital Health Evidence Generation: Multilevel Case Study. *Journal of Medical Internet Research*, *21*(11), e14849. https://doi.org/10.2196/14849

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews A Product from the ESRC Methods Programme Peninsula Medical School, Universities of Exeter and Plymouth*.

Potts, C., Bond, R., Ryan, A., Mulvenna, M., McCauley, C., Laird, E., & Goode, D. (2020). Ecological Momentary Assessment within a Digital Health Intervention for Reminiscence in Persons with Dementia and Caregivers: User Engagement Study. *JMIR MHealth and UHealth*, *8*(7), e17120. https://doi.org/10.2196/17120

Proprietary Association of Great Britain (2016). *Self Care Nation, self care attitudes and behaviours in the UK*. https://www.pagb.co.uk/content/uploads/2016/11/PAGB-SELF-CARE-NATION-REPORT-NOVEMBER-2016-1.pdf

Rahman, Q. A., Janmohamed, T., Pirbaglou, M., Ritvo, P., Heffernan, J. M., Clarke, H., & Katz, J. (2017). Patterns of User Engagement With the Mobile App, Manage My Pain: Results of a Data Mining Investigation. *JMIR MHealth and UHealth*, *5*(7), e96. https://doi.org/10.2196/mhealth.7871

Reinwand, D. A., Crutzen, R., Elfeddali, I., Schneider, F., Schulz, D. N., Smit, E. S., Stanczyk, N. E., Tange, H., Voncken-Brewster, V., Walthouwer, M. J. L., Hoving, C., De Vries, H. (2015). Impact of educational level on study attrition and evaluation of web-based computer-tailored interventions: Results from seven randomized controlled trials. *Journal of Medical Internet Research*, *17*(10), e228. https://doi.org/10.2196/jmir.4941

Richardson, M., Khouja, C., Sutcliffe, K., Hinds, K., Brunton, G., Stansfield, C., & Thomas, J. (2018). *Self-care for minor ailments: Systematic reviews of qualitative and quantitative research.* Department of Health and Social Care Reviews Facility. http://eppi.ioe.ac.uk/

Rubin, G. J., Amlôt, R., Page, L., & Wessely, S. (2009). Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: Cross sectional telephone survey. *BMJ*, *339*(7713), 156. https://doi.org/10.1136/bmj.b2651

Salkovskis, P. M., Rimes, K. A., Warwick, H. M. C., & Clark, D. M. (2002). The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychological Medicine*, *32*(5), 843–853. https://doi.org/10.1017/S0033291702005822

Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement, 66*(6). 1047-1063. https://doi.org/10.1177/0013164406288171

Sheeran, P., Klein, W. M. P., & Rothman, A. J. (2017). Health Behavior Change: Moving from Observation to Intervention. *Annual Review of Psychology*, *68*(1), 573–600. https://doi.org/10.1146/annurev-psych-010416-044007

Shi, J., Lo, B., Wong, H. W., Hollenberg, E., Sanches, M., Abi-Jaoudé, A., Chaim, G., Cleverley, K., Henderson, J., Johnson, A., Levinson, A., Pham, Q., Robb, J., Voineskos, A., & Wiljer, D. (2021). Assessing the usability and user engagement of Thought Spot - A digital mental health help-seeking solution for transition-aged youth. *Internet Interventions*, *24*, 100386. https://doi.org/10.1016/j.invent.2021.100386

Sieverink, F., Kelders, S. M., & Gemert-Pijnen, V. (2017). Clarifying the concept of adherence to ehealth technology: Systematic review on when usage becomes adherence. *Journal of Medical Internet Research, 19*(12), e402. https://doi.org/10.2196/jmir.8578

Sieverink, F., Kelders, S., Poel, M., & van Gemert-Pijnen, L. (2017). Opening the Black Box of Electronic Health: Collecting, Analyzing, and Interpreting Log Data. *JMIR Research Protocols*, *6*(8), e156. https://doi.org/10.2196/resprot.6452

Smith, N., & Liu, S. (2020). A systematic review of the dose-response relationship between usage and outcomes of online physical activity weight-loss interventions. *Internet Interventions*, *22*, 100344. https://doi.org/10.1016/j.invent.2020.100344

Solomon, D., Proudfoot, J., Clarke, J., & Christensen, H. (2015). e-CBT (myCompass), antidepressant medication, and face-to-face psychological treatment for depression in Australia: A cost-effectiveness comparison. *Journal of Medical Internet Research*, *17*(11), e255. https://doi.org/10.2196/jmir.4207

Stassen, G., Grieben, C., Froböse, I., & Schaller, A. (2020). Engagement with a web-based health promotion intervention among vocational school students: A secondary user and usage analysis. *International Journal of Environmental Research and Public Health*, *17*(7). https://doi.org/10.3390/ijerph17072180

Stragier, J., Vandewiele, G., Coppens, P., Ongenae, F., Van den Broeck, W., De Turck, F., & De Marez, L. (2019). Data mining in the development of mHealth apps: assessing in-app navigation through Markov Chain analysis. *Journal of Medical Internet Research, 21*(6), e11934. https://doi.org/10.2196/11934

Szinay, D., Jones, A., Chadborn, T., Brown, J., & Naughton, F. (2020). Influences on the uptake of and engagement with health and well-being smartphone apps: Systematic review. *Journal of Medical Internet Research, 22*(5), e17572. https://doi.org/10.2196/17572

Taki, S., Lymer, S., Russell, C. G., Campbell, K., Laws, R., Ong, K.-L., Elliot, R., & Denney-Wilson, E. (2017). Assessing user engagement of an mHealth intervention: Development and implementation of the Growing Healthy App Engagement Index. *JMIR MHealth and UHealth*, *5*(6), e89. https://doi.org/10.2196/mhealth.7236

Tan, S. S.-L., & Goonawardene, N. (2017). Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. *Journal of Medical Internet Research*, *19*(1), e9. https://doi.org/10.2196/jmir.5729

Trafimow, D., Sheeran, P., Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioural control is a multidimensional construct: Perceived control and perceived difficulty. *British Journal of Social Psychology*, *41*(1), 101–121. https://doi.org/10.1348/014466602165081

Tripepi, G., Jager, K. J., Dekker, F. W., & Zoccali, C. (2010). Selection Bias and Information Bias in Clinical Research. *Nephron Clinical Practice*, *115*(2), 94–99. https://doi.org/10.1159/000312871

Van Der Aalst, W. (2016). *Process Mining: Data Science in Action* (2nd ed.). Springer. ISBN 978-3-662-49851-4

Van Der Mispel, C., Poppe, L., Crombez, G., Verloigne, M., & De Bourdeaudhuij, I. (2017). A self-regulation-based eHealth intervention to promote a healthy lifestyle: Investigating user and website characteristics related to attrition. *Journal of Medical Internet Research*, *19*(7). https://doi.org/10.2196/jmir.7277

Wallston, K. A., Strudler Wallston, B., & DeVellis, R. (1978). Development of the Multidimensional

Health Locus of Control (MHLC) Scales. *Health Education Monographs*, *6*(1), 160–170. https://doi.org/10.1177/109019817800600107

Webb, T. L., Joseph, J., Yardley, L., & Michie, S. (2010). 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*, *12*(1), e4. https://doi.org/10.2196/jmir.1376

West, R., Michie, S., Rubin, G. J., & Amlôt, R. (2020, May 1). Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nature Human Behaviour, 4*, 451-459. https://doi.org/10.1038/s41562-020-0887-9

Williams, S., Yardley, L., Weal, M., & Willis, G. (2010, April 14-16). *Introduction to LifeGuide: Open-source software for creating online interventions for health care, health promotion, and training*. Proceedings of Med-e-Tel 2010, Grimbergen, Belgium. https://www.isfteh.org/files/media/Electronic_Proceedings_of_Med-e-Tel_2010.pdf

World Health Organization (2009). *Hand Hygiene: Why, How & When?* https://www.who.int/gpsc/5may/Hand_Hygiene_Why_How_and_When_Brochure.pdf.

World Health Organization (2010). *What is a pandemic?* https://www.who.int/csr/disease/swineflu/frequently_asked_questions/pandemic/en/.

Wu, J. T., Riley, S., Fraser, C., & Leung, G. M. (2006). Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Medicine*, *3*(9), 1532–1540. https://doi.org/10.1371/journal.pmed.0030361

Xie, L. F., Itzkovitz, A., Roy-Fleming, A., Da Costa, D., & Brazeau, A. S. (2020). Understanding self-guided web-based educational interventions for patients with chronic health conditions: Systematic review of intervention features and adherence. *Journal of Medical Internet Research*. JMIR Publications Inc. https://doi.org/10.2196/18355

Yardley, L., Miller, S., Teasdale, E., & Little, P. (2011). Using mixed methods to design a web-based behavioural intervention to reduce transmission of colds and flu. *Journal of Health Psychology*, *16*(2). https://doi.org/10.1177/1359105310377538

Yardley, Lucy, Joseph, J., Michie, S., Weal, M., Wills, G., & Little, P. (2010). Evaluation of a web-based intervention providing tailored advice for self-management of minor respiratory symptoms: Exploratory randomized controlled trial. *Journal of Medical Internet Research*, *12*(4). https://doi.org/10.2196/jmir.1599

Yardley, L., Miller, S., Schlotz, W., & Little, P. (2011). Evaluation of a Web-based intervention to promote hand hygiene: exploratory randomized controlled trial. *Journal of Medical Internet Research*, *13*(4), e107. https://doi.org/10.2196/jmir.1963

Yardley, L., Morrison, L., Bradbury, K., & Muller, I. (2015). The person-based approach to intervention development: application to digital health-related behavior change interventions. *Journal of Medical Internet Research*, *17*(1), e30. https://doi.org/10.2196/jmir.4055

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F., & Blandford, A. (2016). Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. *American Journal of Preventive Medicine*, *51*(5), 833–842. https://doi.org/10.1016/j.amepre.2016.06.015