Designing evaluation plans for mHealth interventions: A case study on *The Weal Life* app

Patrícia Isabel Amaro

Dissertation written under the supervision of Susana Frazão
Pinheiro

Abstract

Title

Designing evaluation plans for mHealth interventions: A case study on *The Weal Life* app

Author

Patrícia Isabel Amaro

As an applied research project, the purpose of this dissertation is to develop a suitable framework for health-tech companies to evaluate and monitor their mobile-based health technologies' efficacy and usability, whilst defining features and maximizing performance. Indeed, since no officially recognized standards exist for mobile health (mHealth) apps evaluation, despite the fact they can produce health or behavioral benefits, this crucial assessment is often overlooked by companies and poorly integrated into planning.

This research illustrates four mHealth evaluation frameworks and describes the assessment plan that has been developed for *The Weal Life* app*,* a socially connected app targeted to caregivers and circle of support.

Ultimately, and this is our main finding, we identified two necessary conditions for health-tech companies to maximize their mHealth apps evaluation and performance. The first is to embrace a strategic approach that aligns both the mobile health intervention and app development goals, and the second is to assess the impact of these technologies. Indeed, we show that impact measurement process plays the role of "information systems" allowing companies not only to quantify the impact of their products, but also to monitor and improve their performance constantly.

The detail of *The Weal Life* assessment plan will ensure the findings add to evidence and have broad relevance to healthcare professionals and researchers in general.

Key Words: caregiving, evaluation methods, mHealth technologies, mHealth app, evidence-based practice

Resumo

Título

Desenvolver planos de avaliação para intervenções mHealth: Um estudo de caso sobre *The Weal Life app*

Autor

Patrícia Isabel Amaro

Enquanto projeto de pesquisa, o objetivo desta dissertação é desenvolver diretrizes adequadas para as empresas de tecnologias da saúde avaliarem e monitorizarem a eficácia e usabilidade das suas tecnologias móveis, paralelamente definindo especificidades e aumentando o desempenho. Com efeito, uma vez que não existem linhas de orientação reconhecidas no que toca à avaliação de aplicações mHealth, apesar das mesmas poderem ser benéficas para a saúde e comportamento, este estudo crucial é amiúde desprezado por empresas e pobremente executado.

Esta pesquisa ilustra quatro formas de avaliação mHealth e descreve o plano de teste que foi desenvolvido para a app *The Weal Life*, uma aplicação social orientada para os cuidadores e círculo de apoio.

Em síntese, identificámos duas condições necessárias para que empresas tecnológicas na área de saúde possam maximizar a avaliação e performance das suas aplicações mHealth. A primeira consiste em aplicar uma abordagem estratégica que alinhe ambos os objetivos da intervenção tecnológica e do desenvolvimento da aplicação; a segunda, e mais importante, é avaliar corretamente o impacto destas mesmas tecnologias na população-alvo e sociedade em geral. De facto, concluímos que a avaliação funcionará como um "sistema de informação" atualizado, permitindo às empresas quantificar, monitorizar e melhorar a performance dos seus produtos.

O detalhe do plano de avaliação da *The Weal Life* vai fundamentar os resultados postos em evidência e ser relevante para os profissionais de saúde e investigadores em geral.

Palavras-chave: cuidar, métodos de avaliação, tecnologias mHealth, mHealth app, prática com base em provas

Acknowledgements

I would like to thank my supervisors Professor Susana Frazão Pinheiro and Professor Pedro Raposo for the freedom and support provided through the learning process of this dissertation. Furthermore, I would like to thank Keely Stevenson and Josephine Lai for the remarkable opportunity to work within a Startup, as well for the useful comments, remarks and engagement provided along the way.

I would like to express my gratitude to my loved ones, who have guided me throughout the process by keeping me harmonious. Forever grateful for your love.

Table of Contents

List of Tables and Figures

Table 1: Evaluation Schema: Collaborative, Adaptive and Interactive Technology

Table 2: mHealth Evidence Reporting and Assessment (mERA) Guidelines, including mHealth Essential Criteria

Table 3: Evaluation Schema: Collaborative, Adaptive, and interactive Technology

Figure 1: mHealth Device Intervention Classification

Figure 2: The BIT Model

Figure 3: BIT Model in the Context of TIPs

List of Acronyms

mHealth – Mobile Health

FDA – Food and Drug Administration

WHO – World Health Organization

MARS – Mobile Application Rating Scale

RCT – Randomized Controlled Trial

BIT – Behavioral Intervention Technology

TIPs – Trial of Intervention Principles

mTERG - mHealth Technical Evidence Review Group

mERA - mHealth evidence reporting and assessment

1. Introduction

In the beginning of 2015, there were over 259,000 health and fitness related apps available in the two major app stores (i.e. Apple and Android) (Research2Guidance, mHealth Economics Research Program, 2016). In a socioeconomically and geographically diverse sample survey of the US population, Brand., L., and colleagues (2015) found 58,23% of mobile phone owners to have downloaded a health-related app. Yet, despite the proliferation and its acclaimed properties, there remains a lack of evidence of the efficiency and impact these mobile health interventions can have in terms of health outcomes and behavioral change.

As evidence and knowledge are a crucial part of health promotion and implementation, the present research discusses the need for mHealth interventions to be evaluated within a comprehensive framework. Describing its implementation process in a socially connected mHealth app targeted to caregivers - *The Weal Life* app – this dissertation aims to answer the following research questions

1. *Synthetize why health-tech companies should evaluate their mHealth app technologies*
2. *Evaluate the importance and need to adopt a comprehensive approach to mHealth apps and to undertake impact measurement*
3. *Examine how mHealth Apps should be measured*

As a foundation for *The Weal Life* app assessment plan, we'll firstly conduct a thorough review on mHealth technologies, evidence and existing frameworks for its evaluation. Secondly we'll leverage this knowledge in order to build a holistic structure for mobile based health interventions evaluation and monitoring. Thirdly, we'll draw conclusions, recommend a practical approach to be applied in classes, acknowledge limitations and recommend future research.

The Weal Life is a health-tech startup based in San Francisco, California, founded by Keely Stevenson, with the purpose to help family caregivers to cope with the main caregiving challenges and responsibilities. Finally, our main goal is to provide *The Weal Life* a reliable, multidimensional tool for evaluating its mHealth technology, whilst ensuring its intervention to deliver real value to the end users, thus facilitating the caregiving experience. Given the general nature of the assessment plan developed in this research, we hypothesize it to be applicable to a wide range of mHealth apps.

2. Methodology

In order to address the research questions, a qualitative approach in the form of a single case study will be used.

In the context of the collaboration with *The Weal Life* we developed a tailored "Assessment Plan" into the case study, explaining the necessary methodology to quantify and qualify its mHealth app technology benefits and externalities – *The Weal Life App*. Throughout the entire process, we attended weekly meetings with *The Weal Life* founders (Keely Stevenson and Josephine Lai) in order to identify main challenges, concerns and objectives in evaluating their technology. In a jointly effort we developed the final framework. For the purpose, both primary (i.e. company records and weekly meetings with *The Weal Life* team) and secondary data were collected and analyzed. In addition, we did a single interview with the Founder and CEO of Tyze[1], Vickie Cammake. Vickie Cammake is a known and acknowledge entrepreneur in the mHealth US industry. Having found her own mHealth tech company, Vickie widened our perspectives by highlighting the need for mobile based health interventions to be holistically evaluated.

As a starting point, research explored the current state of mHealth adoption and development. Secondly we identified the main challenges in evaluating and generating evidence-based knowledge of mHealth interventions' efficacy. Finally, building on four developed models for socially connected and mHealth apps evaluation, we mapped together all the information into a final comprehensive assessment plan for *The Weal Life* to properly evaluate and maximize its mHealth technology performance. Throughout the entire process,

3. Literature Review

3.1 mHealth: Supporting Health through Technology

"Mobile technologies are mobile and popular, such that many people carry their mobile phone with them wherever they go. This allows temporal synchronization of the

---

[1] Tyze is an online tool that helps people care for others (www.tyze.com)

intervention delivery and it allows the intervention to claim's people's attention when it's most relevant."

<div align="right">Free, C., et al., 2013</div>

Promising to improve healthcare delivery, efficiency and affordability, mobile health technology or simply mHealth proposes to deliver healthcare anytime and anywhere, surpassing geographical, temporal and even organizational barriers (Silva, B., et al., 2015). For the healthcare professional, a mobile device has become an integral part of delivering care, whilst patients have taken an active role in managing their health.

As a new and emergent field, mHealth lacks a standardized definition. Laxminarayan, S., & Istepanian, S., (2000) defined mHealth for the first time as "unwired e-med". In 2011, the Global Observatory for eHealth at the World Health Organization described it as a "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" involving "the use and capitalization on a mobile phone's core utility of voice and SMS as well as more complex functionalities and applications including GPRS, third and fourth generation mobile communications (3G and 4G systems), GPS and Bluetooth technology". In a more general approach, Istepanian, R., and colleagues, (2006) defined mHealth as "the emerging mobile communications and network technologies for healthcare systems".

As it stands, mHealth is a broader term for a myriad of emerging technologies that leverage the scope and speed of mobile networks in order to improve healthcare delivery. In healthcare, the use of mHealth interventions have the potential to improve health outcomes and reduce the cost of care by allowing efficient communication and the simplification of care (Silva, B., et al., 2015). These technologies can support continuous health monitoring, encourage health behaviors, support patient chronic disease self-management, enhance provider's knowledge, reduce the number of healthcare visits, and provide personalized, on-demand interventions (Kumar, S., et al., 2014). On the other hand, concerns regarding safety, privacy of medical information and lack of knowledge also arise.

3.2 The Rise of mHealth Apps: "There's an App for That"

With the rising of mHealth and Smartphone technology, health-related mobile applications[2] (apps) have become highly popularized. Since the beginning of 2015, almost 100,000 apps have been added to the major stores (i.e. Apple store and Android app store), amounting to 259,000 mHealth apps currently available (Research2Guidance, mHealth Economics Research Program, 2016). In a socioeconomically and geographically diverse sample survey of the US population, Brand., L., and colleagues (2015) found 58,23% of mobile phone owners to have downloaded a health-related app. Because most Americans own a smartphone, mobile technologies are a mean to impact health information dissemination and health promotion.

Available apps cover a range of healthcare topics including, wellness & lifestyle, dieting, disease management, fitness and public health. Across the patient spectrum, mHealth apps will provide an opportunity to collect and deliver tailored health information, whilst improving disease self-management and health behavior over time. These can be divided into two broad categories: (1) those specifically focused on disease management interventions (e.g. diabetes monitoring), and (2) those targeted to health behavior change (e.g. smoke cessation programs) (Free, C., et al., 2013).

The widespread use of this technology will have the potential to move healthcare from episodic to continuous care through constant innovation (Kratzke, C. & Cox, C., 2012). Compared to stationary intervention programs, mHealth apps have the potential to be highly interactive and include real-time communication with individuals[3], thus allowing the patients to make decisions and take immediate actions. Using mHealth monitoring, physicians can send and receive patient information in real time using smartphones for motivation, patient adherence, decision support or behavior modification (Fukuoka, Y., et al., 2011; Vervloet, M., et al., 2011). This may improve the patient-provider relationship and support. Through mobile-based technologies, medical references may be accessed in resource limited areas and remote locations (Chang, Y., et al., 2012). However, these "medical apps" also raise several and complex questions regarding

---

[2] Mobile applications or apps, are dowloadable software products that run in mobile devices (Kratzke C. & Cox, C., 2012)

[3] This concept of reaching people throughout their daily life is described as na Ecological Monetary Intervention (EMI) (White, B., et al., 2016)

security, reliability, efficiency and quality of service (Silva, B., et al., 2015). As healthcare providers may consider the integration of smartphones and mHealth apps as a part of the treatment process, it is important to understand how the use of these technologies can leverage health interventions.

3.2.1 Do mHealth Apps Really Provide Benefits?

While more than a third of US physicians claim to have recommended health-related apps to their patients, (Manhattan Research's Taking the Pulse, 2014) there remains uncertainty in doing it so. The lack of evidence for mHealth apps' effectiveness in improving health outcomes and/or changing health behavior continues to be a barrier to the widespread adoption of apps in the healthcare sector.

Whilst systematic reviews are developed, the majority focus on whether apps for a specific condition or health behavior are rooted in evidence-based strategies or theoretical models of behavior change (Boudreaux, E., et al., 2014). Examples include reviews of mHealth apps for cancer prevention and management (Bender, L., et al., 2013), weight loss (Pagoto, S., et al., 2013; Breton R., et al., 2011), mental health (Donker, T., et al., 2013), chronic medical conditions (Martinez-Perez, B., et al., 2013; El-Gayar, O., et al., 2013), HIV prevention (Muessig, E., et al., 2013), and medication self-management (Bailey, C., 2014).

In a meta-analysis aimed to quantify the effectiveness of mobile technology-based interventions delivered to healthcare consumers for health behavior change and disease management, researchers found little or no evidence regarding the benefits of these interventions (Free, C., et al., 2013). From 59 study trials, Free, C., and colleagues report only two mHealth app technologies to be considered for inclusion in healthcare services: (1) a mobile phone short message service (that uses SMS to maintain contact, monitor and respond to medication issues with patients prescribed anti-retrovirals) in Kenya, that showed an increased adherence to anti-retrovirals and clinical important reductions in the viral load (Lester, T., et al., 2010), and (2) an automated text message intervention smoking cessation support that more than doubled biochemically confirmed smoking cessation in the UK (Free, C., et al., 2009; Free, C., et al., 2011). To date, there is insufficient evidence of beneficial effects on clinical outcomes to warrant

implementation of these interventions in other areas of self-management disease or health behavior change (Free, C., et al., 2013).

Lack of information, the heterogeneity of study designs and risk of bias make it difficult to conduct rigorous reviews. Paradoxically, as greater healthcare stakeholders' involvement is required to improve the development of evidence-based apps, the fast-paced growth of the health-related app market outpaces their ability to develop guidance for accuracy, efficacy and security. Currently, evidence is sparse for the efficiency and cost effectiveness of this technology. This is considered to be the major barrier to the implementation of mHealth apps in the healthcare sector (World Health Organization, 2011) and highlights the need for comprehensive evaluation plans for such interventions.

3.3 Evaluating mHealth Apps: A Condition to Leverage Externalities

"Effective evaluation is not an 'event' that occurs at the end of a project, but is an ongoing process which helps decision makers better understand the project; how it is impacting participants, partner agencies and the community; and how it is being influenced by both internal and external factors"

WK Kellogg Foundation, (2004)

mHealth apps are thriving, but before healthcare providers or organizations can recommend an app, they need to be confident the technology will be user-friendly and helpful for the patients they serve (Boudreaux, E., et al., 2014). Although these technologies may seem appealing, research is needed to assess when, where and for whom mHealth devices are efficient (Kumar, S., et al., 2014). Recommending an app to a patient may have serious consequences if its content is inaccurate or if the app is ineffective or even harmful (Boudreaux, E., et al., 2014). In a recent study, Wolf, A., and colleagues (2013) found 3 out of 4 smartphone apps in assessing melanoma risk misclassified 30% or more of melanomas. Reliance on these apps, (which are not regulated) have the potential to delay the diagnosis and treatment for a condition in which early detection has a significant impact on survival rates. However, the

development of smartphone apps has not traditionally included healthcare professionals input and no standardized evaluation tools are available for health specialists to review content quality of health-related apps (Kratzke, C., & Cox, C., 2012). As with any other procedure, before mobile-based health assessment methods can be recommended, their reliability must be established (Barnhart, X., at al., 2007) and guidance on how to judge their commercial value must be acknowledged (Boudreaux, E., et al., 2014).

To date, little or no quality control exist to ensure mHealth quality and validity. The Food and Drug Administration (FDA) does not regulate most mHealth apps and further suggests that it does not intend to expand regulation to non-medical devices (United States Food and Drug Administration, 2014). To clarify the question if smartphones can be classified as medical devices (thus regulated by the FDA), the FDA proposed to amend its regulations governing classification of medical devices, releasing in February 2015, a new guidance for mobile apps. The FDA states "when the intended use of a mobile app is for diagnosis of disease or other conditions, or the cure, mitigation, treatment, or prevention of disease, or is intended to affect the structure of any function of the body of man, the mobile app is a medical device" (U.D. of Health, 2015)

Consequently, as these initiatives proliferate, academic researchers develop several models and tools to guide its evaluation (White, B., et al., 2016). The following four approaches balance the efficiency of the research process against the need to evaluate the accuracy and validity of a mHealth app.

3.3.1 Collaborative Adaptive Interactive Technology

In 2009, Grady, L., and colleagues developed the "Collaborative Adaptive Interactive Technology" dynamic framework. Whilst developed for web-based interventions, the framework is adjustable and particularly significant for mHealth applications that connect people and groups. This evaluation tool is particularly relevant for our project as it aims to evaluate the end user (not specifically the patient) and all the people involved (i.e. a socially connected app) in using or developing the mHealth app. As stated in the original work, "people contains parameters related to the individuals who are involved in using or developing the technology, or may be affected by the implementation of the technology" (Grady, L., et al., 2015)

In their research, the authors define "collaborative, adaptive and interactive technologies" as those that (1) facilitate collaboration between users, (2) adapt content according to users' needs and (3) enable user interaction (O'Grady, L., et al., 2009). Additionally, the concept encompasses five major aspects of health applications: (1) social networking, (2) participation, (3) apomediation, (4) collaboration, and (5) openness.

The framework organizes formative (i.e. development and laboratory testing), summative (i.e. efficacy and goal achievement), and outcomes (i.e. impact assessment) indicators into five major areas:

1 – People: end users and stakeholders

2 – Content: information and web content

3 – Technology: technology used to create and develop the website

4 – Computer-Mediated Interaction: assessment of user interactions with and via the technology

5 – Health Systems Integration: how the website impacts the broader health system, processes and the society

Table 1 – Evaluation schema: collaborative, adaptive and interactive technology (O'Grady, L., and colleagues, 2009)

	Formative *Development and laboratory testing*	Summative *Efficacy and goal achievement*	Outcome *Impact assessment*
People	• Identification of Stakeholders' characteristics and needs • Assessment of stakeholders' interests	• End users' traits (computer proficiency, health literacy, cognitive style, affective traits) • End users' perspectives (intention to use, satisfaction, motivation to use)	• Patient outcomes • Impact on interpersonal relationships (patient-physician, caregiver-patient)
Content	• Quality and credibility • Utility (completeness, relevance, understandability)	• Quality and credibility • Subjective utility • Level of personalization	• Content produced (form and nature) • Positioning of user-generated content
Technology	• System robustness (performance,	• Usage statistics (hits, visitors, browsers,	• Dynamic evolution • Collaborative

	functionality and features, security and privacy) • System interoperability • Privacy	errors) • System reliability • Speed • Positioning within current technology • Standards compliance	development models (E.g. Open source)
Computer-Mediated Interaction	• Usability • Accessibility • Sociability • Interactivity • Information architecture	• End users' perspectives on usability and accessibility • Demonstrated sociability and interactivity • Collaboration • Findability	• Community development • Evolution of collaboration
Health Systems Integration	• Definition of evaluation metrics and processes • Ethics/liability	• Administration (services utilization, care coordination, patient safety)	• Public impact • Cost-effectiveness • Intended effect (appropriateness and effectiveness)

3.3.2 Mobile Application Rating Scale

The Mobile Application Rating Scale (MARS) was developed by Stoyanov, S., and Psych, M., in 2015. This recently developed tool evaluates the quality of a mHealth App by addressing five broad categories of criteria:

1- Engagement: entertainment, interest, customization, interactivity and target group fitting
2- Functionality: performance, ease of use, navigation and gestural design
3- Aesthetics: layout, graphics and visual appeal
4- Information Quality: accuracy of the app description, goals, quality of information, quantity of information, visual information, credibility, and evidence base
5- Subjective Quality Scale: users' opinion (e.g. Would you recommend this app to people who might benefit from it?)

Each item uses a 5-point scale (*1 – Inadequate, 2 – Poor, 3 – Acceptable, 4 – Good, 5 - Excellent*) and it is scored by calculating the mean values of the four objective subscales (engagement, functionality, aesthetics and information) and an overall mean quality

total score. This tool can be easily incorporated into an app development process, identifying areas of weakness and strength, or used to rate an already existing app.

3.3.3 Trial of Intervention Principles

In 2015, Mohr, D., and colleagues suggested traditional randomized controlled trials (RCTs) methodologies not to be ideal for evaluating "Behavioral Intervention Technologies" (BITs), (i.e. eHealth and mHealth interventions expected to be modified frequently), as "RCTs usually test early versions of BITs that often have to undergo revisions prior to implementation" (Mohr, D., et al., 2015) and researchers often refuse to improve the BIT during the RCT (based on new trial information, such as bugs and errors).

In their work, the "Trial of Intervention Principles" (TIPs) analysis focus on assessing evaluation and intervention principles, allowing for ongoing quality improvements to the technology, (as opposed to a static traditional version of RCTs).

Mohr, D., and colleagues (2015) start by recommending a specific BIT framework (See Figure 1) including two broader categories: (1) a conceptual action level (fixed) which reflects the intentions of the intervention (i.e. why the BIT exists), and (2) an
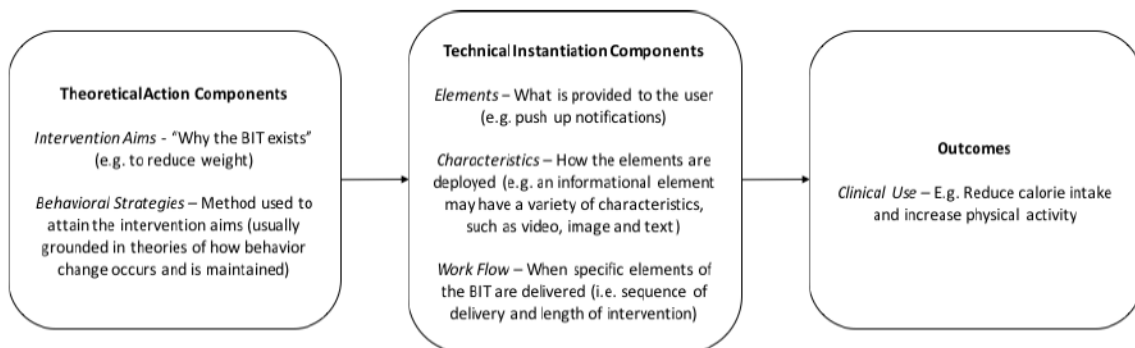

Theoretical Action Components

Intervention Aims - "Why the BIT exists" (e.g. to reduce weight)

Behavioral Strategies – Method used to attain the intervention aims (usually grounded in theories of how behavior change occurs and is maintained)

Technical Instantiation Components

Elements – What is provided to the user (e.g. push up notifications)

Characteristics – How the elements are deployed (e.g. an informational element may have a variety of characteristics, such as video, image and text)

Work Flow – When specific elements of the BIT are delivered (i.e. sequence of delivery and length of intervention)

Outcomes

Clinical Use – E.g. Reduce calorie intake and increase physical activity

Figure 2 – The BIT Model (Mohr, D., and colleagues, 2015)

instantiation category (dynamic, i.e. that can be adjusted during the trial) that reflects the technological implementation (i.e. what is delivered, when and how).

Figure 2 places the BIT model into the context of TIPs. Aspects included in the dotted box should not be changed during the course of a RCT. Instantiation components (if not a part of the principles) can be adjusted during the trial. The outcomes box represents the measurement of intervention outcomes.
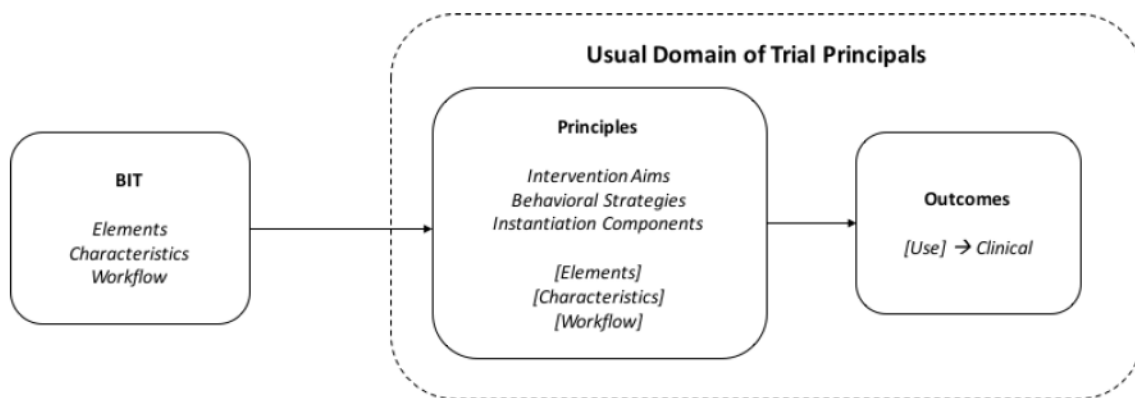
Figure 3 – BIT Model in the context of TIPs (Mohr, D., and colleagues, 2015)

Usually, the decision to change a BIT is made due to a belief, that a specific instantiation component – elements, characteristics or workflow - is producing suboptimal usage or intervention outcomes.

Thus, the TIPs framework integrates adjustments and quality improvements into RCTs, (as long as changes are reported) allowing for ongoing learning and optimization of the tested technology.

3.3.4 World Health Organization Evidence Reporting and Assessment Checklist

In 2016, the WHO mHealth Technical Evidence Review Group (mTERG) developed a mHealth evidence reporting and assessment (mERA) checklist consisting of 16 key criteria (See Table 2) (Agarwal, S., Lefevre, A., et al., 2016). The checklist identifies a minimum set of information needed to outline (1) what the mHealth intervention is (content), (2) where it is being implemented (context), and (3) how it was implemented (technical features). It includes several items such as the mHealth infrastructure, technology platform, user feedback and usability, and it aims to improve transparency in reporting, promoting a critical assessment of mHealth research evidence, and helping to improve the precision of future reporting of research findings (Agarwal, S., Lefevre, A., et al., 2016).

Table 2 – mHealth evidence reporting and assessment (mERA) guidelines, including mHealth essential criteria (Agarwal, S., Lefevre, A., et al., 2016)

Criteria	Item No	Notes
Infrastructure (population level)	1	Availability of infrastructure to support technology operations in the

18

		study location (e.g. physical infrastructure such as electricity and connectivity)
Technology platform interoperability	2	Justification for the technology architecture. This includes a description of the software and hardware
Health information systems (HIS) context	3	Describes how the mHealth intervention can integrate into existing HIS
Intervention delivery	4	Description of the mHealth intervention (e.g. frequency of mobile communication, mode of delivery of intervention, timing and duration)
Intervention content	5	Description of the details of the content (as well as source and any modifications to the content)
Usability/content testing	6	Describes formative research and/or content and/or usability testing with target groups clearly identified
User feedback	7	Describes user feedback or user satisfaction with the intervention (e.g. usability, access, connectivity)
Access of individual participants	8	Mentions barriers or facilitators to the adoption of the intervention among study participants (e.g. economic or social barriers/facilitators)
Cost assessment	9	Presents basic costs assessment of the mHealth intervention from varying perspectives
Adoption inputs/program entry	10	Describe how people are informed about the program (e.g. promotional activities or training if relevant)
Limitations for delivery at scale	11	Cleary presents the mHealth solution limitations for delivery at scale
Contextual adaptability	12	Describes the adaption (or not) of the solution to a different language, population or context. Any tailoring or modification of the intervention that resulted from pilot testing/usability is described
Replicability	13	Detailed intervention to support replicability (e.g. code/flowcharts of algorithms)
Data security	14	Describes the data security procedures/confidentiality protocols
Compliance with national guidelines or regulatory structures	15	Mechanisms used to assure that content or other guidance/information provided by the intervention is in alignment with the existing national/regulatory guidelines
Fidelity of the intervention	16	Describes the strategies employed to assess the fidelity of the intervention (i.e. Was the intervention delivered as planned?)

4. Case Study – *The Weal Life* App Assessment Plan

"Healthcare providers may want to stay informed due to interest in recommending mobile health apps as part of their treatment planning (…). However, guidance on how to judge the validity or worth of commercially available apps is lacking. The decision to recommend an app to a patient can have serious consequences if its content is inaccurate or if the app is ineffective or even harmful."

Boudreaux, E., et al., 2014

4.1 mHealth Apps: Supporting Health through Technology

Mobile devices offer unique opportunities to design tailored solutions and reach populations. For healthcare professionals, a mobile device has become an integral part of delivering care, whilst patients have taken an active role in managing their health.

In 2015 more than half (58,23%) of the mobile phone owners' US population had downloaded a health-related app[4]. In 2016, more than 259,00 mobile health (mHealth) apps were available in the major app stores[5]. Yet, despite its proliferation, there remains lack of evidence and regulation in terms of mHealth efficacy and impact in health outcomes and behavioral change.

The need for more evidence regarding the effectiveness of mHealth apps is commonly acknowledged, yet there are significant challenges in designing and implementing evaluation plans. In particular, the difficulty of trialing an app for the recommended period of time, and the potential for changes in technology or expectations from individuals. As the end-user's usage and engagement with an app is complex, multiple factors (e.g. robustness of the technology, engagement strategies, interaction of the user with the technology, etc.) can impact the success of an intervention. To generate evidence and improve health promotion, a comprehensive (and dynamic) evaluation plan is required.

4.2 *The Weal Life* App

"We use technology to make it easier to care for one another, especially during a health crisis, aging or chronic illness. We value personal relationships by making the experience of caring for our loves ones less overwhelming, more collaborative and more meaningful & joyful"

Keely Stevenson, Founder and CEO of *The Weal Life*

The Weal Life is a socially connected mHealth app designed specifically for helping caregivers to cope with the daily challenges of caring for a loved one. With a focus on

---

[4] Brand., L., and colleagues, 2015
[5] Research2Guidance, mHealth Economics Research Program, 2016

four main functionalities – (1) task creation and scheduling (management tool), (2) chat, (3) gift registry and (4) library (*"get ideas"* feature in the app) - *Weal* recognizes trusted friends and family as a circle of support and makes it easier, more transparent and rewarding to help (See **Exhibit 1** for snapshots of the app).

The app utilizes several engagement techniques including push notifications and humor (*"to make the experience a little lighter, a little more joyful"* – Keely Stevenson, November 2015). *The Weal Life* has an extensive, searchable evidence-based information library (i.e. articles on a range of topics related to a specific disease/condition and ideas on how to help your loved one) and a wish list feature of needs and gifts easily connected to Amazon. Caregivers can create a new team or family members and friends can log in/be invited to an existing one, enabling users to create or assigned themselves to a specific task (e.g. Pick up Kim from chemotherapy at 5pm) and connect with other supporters and share messages.

The Weal Life is the first app we are aware that focus on the patient and support circle informational (i.e. understanding all aspects of the situation, illness, medical/healthcare related resources and life resources) and psychological needs (i.e. feeling supported, hopeful, and a sense of control of the situation).

Leveraging mobile technology to capture the value of relationships, *The Weal Life* intervention aims to increase the emotional, financial and logistical support family members and friends can give to an ill/impaired loved one and its primary caregiver; which *Weal* hypothesizes will increase the perceived social support and in particular reduce the caregiver's burden level.

4.3 *The Weal Life* Assessment Plan

In developing the assessment plan for *The Weal Life* app, there was a focus on planning for ongoing evaluation throughout the app development cycle. This included three stages: (1) formative (while planning the app intervention), (2) summative (during the app implementation) and (3) outcome evaluation (studying the outcomes of the app intervention).

The development of the app included formative evaluation with input from the end-user group (i.e. primary caregivers, care recipients and circle of support), and a user-testing phase.

As the pilot study trial is currently proceeding, the summative and outcome evaluation are still being conducted. The data for these stages will be collected from participants via self-administered questionnaires at recruitment, baseline (month 1) and follow-up (month 6). *The Weal Life* assessment plan is detailed in **Exhibit 2**, and focuses generally on evaluating the delivery of social, emotional, financial and logistical support to caregivers and care recipients through a mHealth app intervention. The inclusive plan is based on the "Collaborative Adaptive Interactive Technology" dynamic framework developed by O'Grady, L., and colleagues (2009), designed to evaluate items for the formative, summative and outcome assessment of *The Weal Life* across five areas – people, content, technology, computer-mediated interaction and health systems integration. The five broad categories of criteria are described below.

People

In the case of *The Weal Life,* "people" refers to end-users (i.e. caregivers, care recipients and circle of support) and stakeholders (i.e. health insurance companies and healthcare professionals). The formative evaluation stage sought to identify only the end-users' needs and preferences. This included a literature review, the creation of personas[6], user-testing (with think-aloud walkthrough[7]) and individual interviews. Stakeholders were not considered important to include at this stage as the company considered them not to have practical insights or enough awareness about the caregiving experience. The end-user group focused on the framing of the app, and on the acceptability of the proposed functionalities and design. Insights about the caregivers' perspectives, individuals' intentions, needs, problems and motivations to use the app were included at all evaluation levels.

Data was gathered by the conclusion of the think-aloud walkthrough studies and summative and outcome evaluation data will be collected via questionnaires regarding

---

[6] Personas are a way to model, summarize and communicate about the end-users' or stakeholders' traits, needs and preferences. Whilst personas do not represent a real individual, they synthetize significant information regarding the target population – See **Exhibit 3**

[7] In think-aloud walkthroughs the end-user is observed using the app while communicating their thoughts. This process can help the developers to highlight issues with navigation or usability.

the caregivers' burden levels (Zarit Burden Interview[8]) and perceived social support (Social Support Index[9]) as part of the pilot study randomized controlled trial (RCT). For the RCT we'll target caregivers with specific characteristics (please see **Exhibit 6** for a detailed explanation of the sample selection procedures). The outcome assessment includes examination of the data gathered at recruitment, baseline (month 1) and follow-up (month 6), including *The Weal Life* mHealth impact on caregivers' burden levels and perceived social support, as well as the completion of the Mobile App Rating Scale[10] (MARS).

Content

The "content" in the case of *The Weal Life* refers to the constantly updated informational content on the "library" section (i.e. *"get ideas"* feature in the app) and the user-generated content in the chat, in the task creation management tool and in the wish list feature.

In the development of the app, formative evaluation sough to test the quality and reliability of the library content through extensive literature reviews, caregivers' perspectives, and healthcare professional insights. Understandability, functionality and usability will be measured in the summative phase via the MARS questionnaire completion.

The Weal Life contains a socially connected chat for caregivers and supporters to interact with each other, a task management tool feature for users to create or assign themselves to a specific task and a wish list feature where users can create and purchase gifts. The outcome assessment will focus on collecting computerized measures in order

---

[8] The Zarit Burden Interview is a self-reported measure developed by Zarit, H., and colleagues in 1985. The total score of the Zarit Burden Interview reflects the overall level of burden perceived by the caregiver within the caregiving episode – **Exhibit 4**

[9] The Social Support Index scale developed by McCubbin, I., Patterson, J., & Glynn, T., in 1982, aims to measure family and friends social support, as well as the amount of support families believe to exist within the community – **Exhibit 5**

[10] The Mobile App Rating Scale (MARS) developed by Stoyanov, S., and Psych, M., in 2015, is a recently developed tool that evaluates the quality of an App by addressing five broad categories of criteria: engagement, functionality, aesthetics, information quality and one subjective quality scale – See **Exhibit 7** for a subsection of the MARS questionnaire

to understand what role do these features play for users and what position do they have in the app, for instance: level of chat activity, total number of people invited to a team Vs. the number of people who actually joined the team, and total number of tasks created Vs. the total number of tasks accepted.

Technology

The software used to develop and run *The Weal Life* app is the "technology". In this category, the formative phase included indicators such as the quality of the system (performance and speed), as well as how it performed with several users. Privacy of users and generated content was an important consideration for *The Weal Life,* including how the data is stored in order to ensure confidentiality.

Summative assessment will involve an examination of usage statistics embedded in the app to answer questions such as, how people are using the app, which features are the most popular and how supporters are engaging with a team. In the outcome phase, we will look at the dynamic of *The Weal Life* evolution over time (e.g. how it adapts to new technologies, new operating system updates, new mHealth applications and social trends).

Computer-mediated technology

"Computer-mediated technology" refers to the evaluation of the end-users' interactions with and via *The Weal Life* interface technology. Formative assessment included the think-aloud walkthroughs that measured the app usability. Additional information will be collected from end-users regarding their perspectives on the functionality of the app through the completion of the MARS questionnaire. The final outcome evaluation will examine content analysis and app-related metrics to explore the "community engagement" within the app (i.e. How *The Weal Life* has supported the creation of a "community" around the caregiver and care recipient, and how users have engaged with each other over time).

Health Systems Integration

The category "health systems integration" refers to the larger healthcare landscape and society, in which *The Weal Life* may be included. In this section, the formative

assessment included the development of a comprehensive framework including the inclusion of ethical and legal concerns. Summative evaluation will measure the impact of *The Weal Life* in caregivers' usage of other health services (such as external websites and/or other mHealth applications, and visits to healthcare professionals in relation to the caregiving experience). The outcome evaluation of *The Weal Life* app will include considerations on its sustainability and how it can be embedded into existing healthcare insurance plans and health services (i.e. the stakeholders). Finally, the RCT pilot study includes a cost-effectiveness analysis of the intervention being analyzed.

4.4 Discussion

As mHealth technologies continue to proliferate, regulators and healthcare professionals will increasingly require the ability to plan and evaluate technological interventions. Challenges persist in developing mHealth applications, including privacy and security concerns, and the need for more evidence-base regarding their efficacy, usability and cost-effectiveness. Will the developed framework be able to continuously monitor and evaluate mobile based health applications? Or will this toolset, once again, be outdated by the fast-paced development of technology?

4.5 Exhibits

Exhibit 1 – *The Weal Life* app snapshots *(The Weal Life internal documentation 2015)*

Figure 1 – Log in or join an existing team
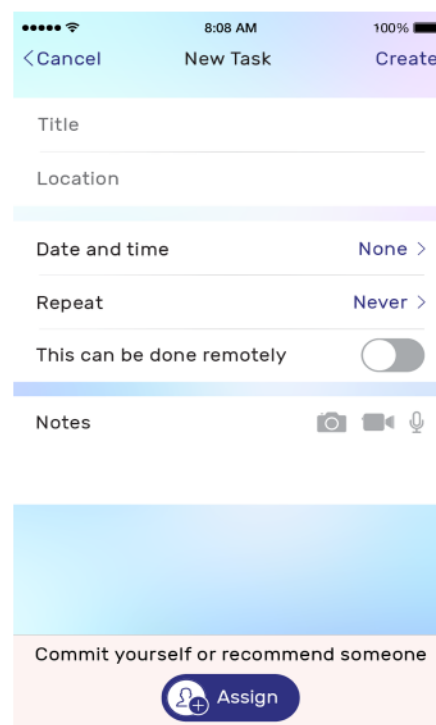
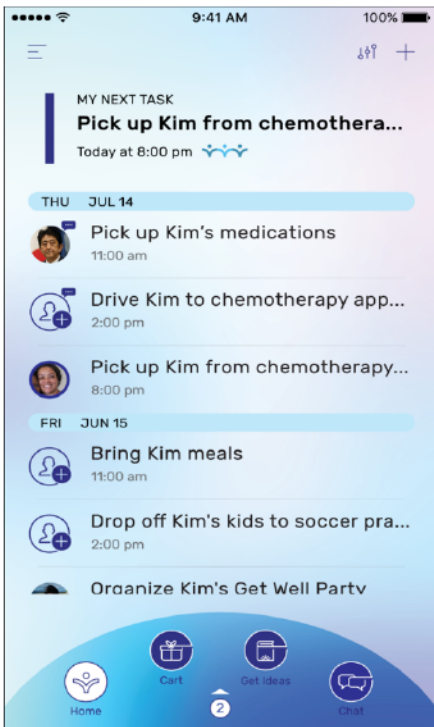**Figure 1** – Task creation feature (how to create a new task)
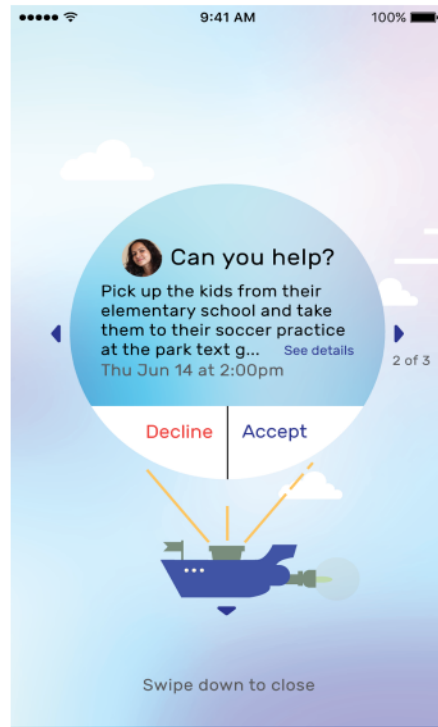
Figure 2 – Task creation/scheduling feature

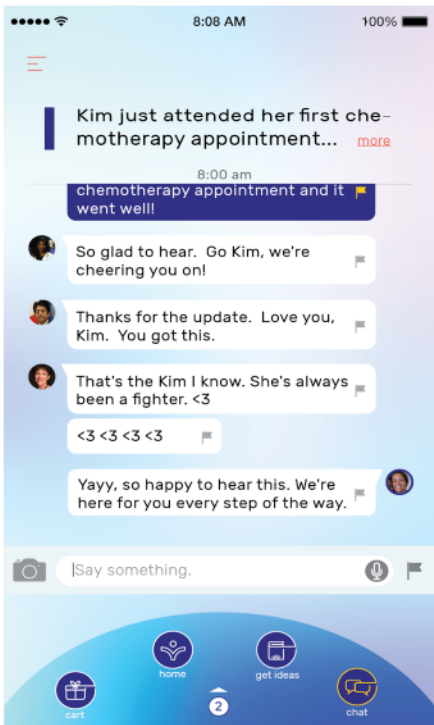**Figure 4** – Engagement technique (make it easier for supporters to take action)

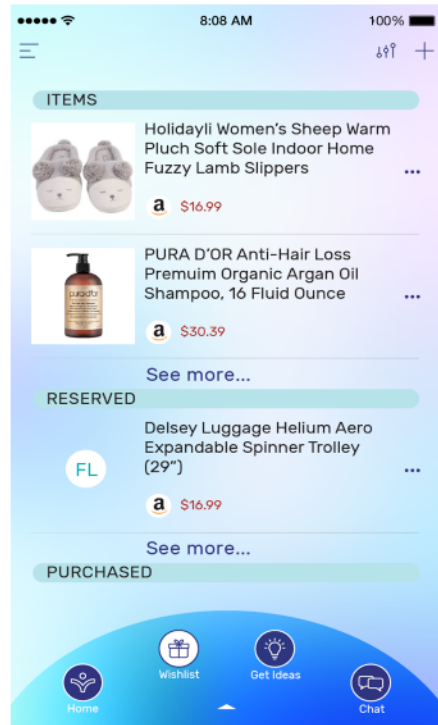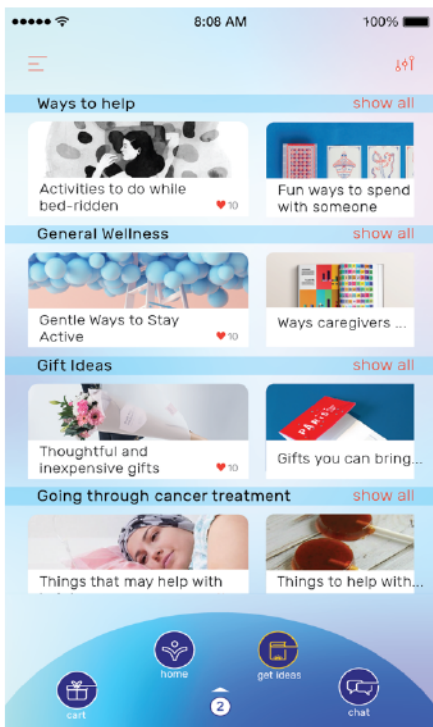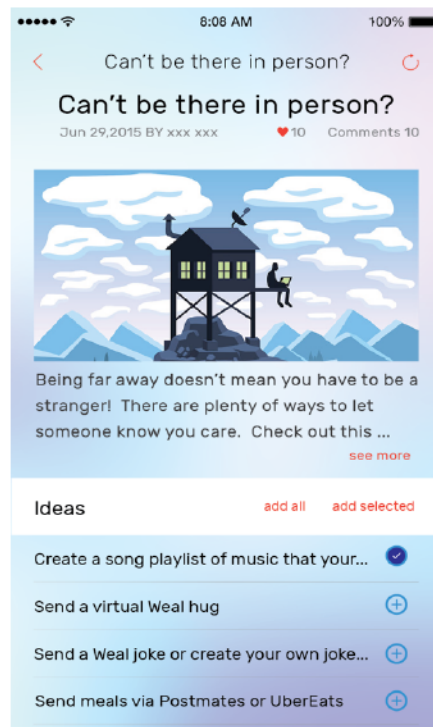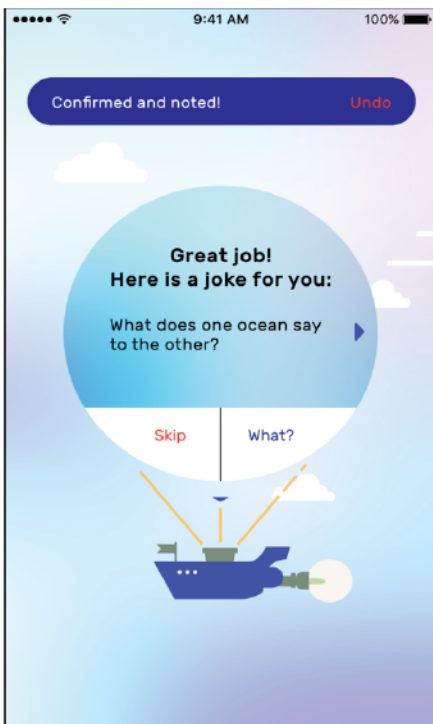**Figure 5** – Chat feature (connect with the team and share messages)

Figure 6 – Wish list feature (supporters can view and reserve items to purchase)

Figure 7 – Library "get ideas" feature (articles on a range of topics)

Figure 8 – Get ideas on how to help and add them to your tasks

Figure 9 – Engagement technique (humor)

Exhibit 2 – *The Weal Life* assessment plan (adapted from O'Grady, L., and colleagues, 2009)

	Formative Evaluation		Summative Evaluation		Outcome Evaluation	
	Indicator	Measurement	Indicator	Measurement	Indicator	Measurement
People *(caregivers, care recipients and circle of support)*	**End users' traits:** - Identification of key characteristics of the target group **End users' perspectives:** - Identification of target group needs, and problems - Identification of target group interests	- Individual Interviews - Creation of Personas - Literature Review - Think-aloud walkthrough	**End users' perspectives:** - Intention and motivation to use the app - Satisfaction	- MARS questionnaire	**Health Outcomes:** - Impact on caregivers' burden levels and perceived social support **Constructs that mediate the caregiver experience:** - Knowledge - Emotional support - Quality of the circle of support - Caregiving efficacy **Impact on interpersonal relationships:** - Caregiver/care recipient relationship	- General linear mix model analysis (differences-in-differences) and correlation between app usage and specific determinants
Content *(library content, user-generated content in the chat and wish list, and creation of tasks)*	**Quality and credibility of the app content:** - Understandability - Relevance - Completeness of content	- Literature review - End users' perspectives - Healthcare professionals insights	- Quality and credibility - Subjective Utility - Engagement techniques - Design	- MARS questionnaire	**Content generated by users:** - Nature (emotional, logistical or financial support) - Positioning of content	- App analytics - Content analysis of user-generated activity
Technology	**System robustness:** - Performance - Loading speed - Privacy of data - Security	- Alpha/Beta testing - Developers team testing	**Usage statistics:** - Number of logins and usage - Patterns of usage - Detected "bugs" **System reliability:** - Speed - Performance	- App analytics - Questionnaires	**Dynamic Evolution:** - App response to software updates - Ongoing nature of the app - App response to social trends	- Response to updates - Recommendation from findings
Computer-Mediated Interaction	**Information architecture:** - Organized to support the user - Sociability	- Alpha/Beta testing - User testing (think-aloud walkthroughs)	**Users' perspectives on:** - Usability - Accessibility - Sociability and collaboration (mechanisms to support community) - Findability (e.g. were users able to find what they were looking for?)	- Questionnaires	**Community development:** - How did the app support the community creation and interaction? **Evidence of collaboration:** - How did the community develop over time? - What was the nature of collaboration?	- App analytics - Content analysis of users comments

Health Systems Integration	- Ethical and legal concerns	- All necessary ethical clearances	Service utilization: - Ways in which the app may be affecting delivery of health services (impact on other services utilization)	- Questionnaires	Public impact: - Effect on the wider community - Effectiveness in reaching health outcome goals - Cost-effectiveness	- Statistics analysis - Questionnaires - Economic analysis

Exhibit 3 – Example of Personas developed for the formative evaluation of the end-user *(The Weal Life internal documentation 2015)*

Exhibit 4 – *Abridged Version of Zarit Burden Interview* used for the outcome evaluation of the end-user *(Zarit, H., and colleagues, 1985)*

All questions are answered as (0) "never", (1) "rarely", (2) "sometimes", (3) "quite frequently", or (4) "nearly always".

"Do you feel..."

1. that because of the time you spend with your relative that you don't have enough time for yourself?

2. stressed between caring for your relative and trying to meet other responsibilities (work/family)?

3. strained when you are around your relative?

4. uncertain about what to do about your relative?

Exhibit 5 – *Social Support Index* used for the outcome evaluation of the end-user *(McCubbin, I., Patterson, J., & Glynn, T., 1982)*

All statements are rated as (1) "strongly disagree", (2) "disagree", (3) "neutral", (4) "agree", or (5) "strongly agree".

1. If I had an emergency, even people I do not know in this community would be willing to help

2. I feel good about myself when I sacrifice and give time and energy to members of my family

3. The things I do for members of my family and they do for me make me feel part of this very important group

4. People here know they can get help from the community if they are in trouble

5. I have friends who let me know they value who I am and what I can do

6. People can depend on each other in this community

7. Members of my family seldom listen to my problems or concerns: I usually feel criticized

8. My friends in the community are a part of my everyday activities

9. There are times when family members do things that make other family members unhappy

10. I need to be very careful how I much I do for my friends because they take advantage of me

11. Living in this community gives me a secure feeling

12. The members of my family make an effort to show their love and affection for me

13. There is a feeling in this community that people should not get too friendly with each other

14. This is not a very good community to bring children up in

15. I feel secure that I am as important to my friends as they are to me

16. I have some very close friends outside the family who I know really care for me and love me

17. Members of my family do not seem to understand me; I feel taken for granted

Exhibit 6 – Sample Selection and Criteria

For the purpose of the study, The Weal Life will collect a convenience sample[11] of the US primary caregiver population[12]. Inclusion criteria comprises: (1) being primary caretaker of an ill relative/friend/neighbor, (2) younger than 60 years old, (3) English speaker, (4) smartphone owner, (5) living in San Francisco Area. In order to have valid results, it will be important to include in the analysis the app usability by the end-users (i.e. monitor app analytics, such as number of logins per week). Through the completion of the MARS questionnaire throughout the evaluation plan we'll also be able to understand if a caregiver doesn't find the app useful or resourceful.

Whereas there isn't an ideal sample size, it is important for the sample to be similar/representative of the U.S. general caregiver population (See Table 1).

Caregivers' Characteristics	General US (2015)
Caregiver's gender:	
• Female	75%
Care recipient's gender	65%

---

[11] Given resources constraints (i.e. size of the company and stage of development) a convenience sample will be used. *Convenience sampling* is a non-probability sampling method that relies on data collection from the population members who are conveniently available to participate in the research.

[12] *Primary Caregiver* refers to the individual who assumes the major responsibility of caring for and making treatment decisions for the care recipient.

- Female

Distribution of caregiver's age:

• 18-34	24%
• 35-49	23%
• 50-64	34%
• 65-74	12%
• 75 or older	7%

Distribution of care recipient's age:

• 18-49	14%
• 50-75	39%
• 75 or older	47%

Individual caregiving prevalence by ethnicity:

• White (Non-Hispanic)	
• African American	16,9%
• Asian American	20,3%
• Hispanic	19,7%
	21%

Care recipient relation to caregiver:

• Parent	42%
• Friend	15%
• Child	14%
• Parent-in-law	7%
• Grandparent or grandparent-in-law	7%

Duration of care:

• Less than 5 years	75%
• 5 years or more	24%

Table 1: Caregivers' Characteristics and Responsibilities in 2015 (AARP Institute, 2015)

Over a 3-month period, *The Weal Life* will approach family caregivers within its own network of contacts. For those who decide to enroll, an informative email will be sent. Through a link embed in the email, individuals will have access to a descriptive survey – in order to collect participants' demographic information. Participants who don't complete the initial survey will be excluded.

Exhibit 7– *Mobile App Rating Scale (MARS)* questionnaire used in several evaluation stages *(Stoyanov, S., and Psych, M., in 2015)*

All items are rated on a 5-point Likert scale from "1 - Inadequate" to "5 - Excellent".

1. Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment? (e.g. through gamification)?
2. Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its components in an interesting way?
3. Customization: Does it provide/retain all necessary setting/preferences for app features (e.g. sound, content, notifications, etc.)?
4. Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: These functions need to be customizable and not overwhelming in order to be perfect.
5. Target Group: Is the app content (visual information, language, design) appropriate for your target audience?

5. Teaching Notes

5.1 Learning Objectives

The Weal Life App Assessment Plan case study, aims to help students understand how technological companies can measure, monitor and maximize the impact of mHealth

app interventions, whilst raising awareness for the importance of providing clear-based evidence about mHealth products. Additionally, students attending technological, innovation or entrepreneurial seminars can become familiar with a successfully developed mHealth app and understand the relevant factors to be considered when building a health-tech solution.

The proposed learning objectives of this case are:

1. To understand the importance of properly evaluate a mHealth app intervention;

2. To highlight the importance of properly developed indicators, metrics and outcome evaluation;

3. To understand the relevant components and phases to be considered while developing a health-tech solution;

4. To become aware of an existing dynamic framework for mHealth solutions evaluation;

5. To be able to apply the same comprehensive evaluation plan to a project developed by themselves or proposed by a Professor;

6. To become familiar with an existing mHealth solution for caregivers;

5 2 Teaching Questions

In order to support a brainstorming and practical session on mHealth app evaluation plans, the following three teaching questions (TQs) are recommended.

TQ1: Why do mHealth app developers need to incorporate comprehensive evaluation plans into their products?

TQ2: Which are the key factors that mHealth app developers have to consider in order to properly assess their technology?

TQ3 (if applicable): Use the previous framework and insights to develop your own app assessment plan.

Through their own reading, students should individually reflect on *The Weal Life Assessment Plan* case study. They are expected to realize the need to implement

comprehensive evaluation plans into mHealth technology, whilst understanding its basic structure and framework.

The professor can actively create groups of 4-5 students, and guide the practical and brainstorming session with the help of the teaching questions and learning objectives. If applicable (for instance, in a Lean Entrepreneurship Project Class) students can use the presented evaluation framework to create an assessment plan to evaluate their own projects.

5 3 Discussion

The following answers will provide some insights in how to properly explore *The Weal Life Assessment Plan* case study. The proposed teaching questions are exclusively based on the case study and *The Weal Life* example.

TQ1: Why do mHealth app developers need to incorporate comprehensive evaluation plans into their products?

This answer can be found on the first (4.1) and final (4.4) chapters. mHealth app developers need to incorporate comprehensive evaluation plans into their products because:

- mHealth apps are evolving technologies inserted in a dynamic and rapidly changing environment (technological and consumers' expectations constantly change and adapt to new trends and needs). Locking down an intervention during a static trial decreases the likelihood that useful information will come from the experiment (Mohr, D., et al., 2015);
- As evolving (rather than static) technologies, mHealth apps need to be evaluated through dynamic and comprehensive plans - able to adjust and constantly improve. Integrating ongoing quality improvement into assessment plans can spur innovation, allowing for the mHealth technology to develop over time (Mohr, D., et al., 2015);
- As a proliferating technology, healthcare providers and regulators will increasingly require the ability to properly evaluate mHealth technologies efficacy and outcomes, in order to include them in their healthcare plans and

recommendations. "We need to determine whether something is effective before it can be of value" (O'Grady, L., and colleagues, 2009);

- As a complex construct, the success of a mHealth intervention is dependent on multiple factors (such as: robustness of the technology, engagement strategies, interaction of the user with the technology, and design);
- There is little evidence-base regarding the efficacy, usability, cost effectiveness and general health outcomes of mHealth app interventions;

Additionally, students should brainstorm other ideas that have not been considered, but that could be relevant. For instance:

- Reliance on mHealth interventions, which are not subject to regulatory oversight, can delay the diagnosis and even harm patients. In 2013, researchers from the University of Pittsburg Medical Center (Wolf, J., Moreau, J., et al., 2013) found three out of four skin cancer diagnosis apps to miss at least 1 in 3 melanomas. mHealth interventions carry the risk of causing more harm than good (O'Grady, L., and colleagues, 2009);
- Evidence about the efficacy of mHealth applications is an emerging field and appropriate evaluation skills are needed (White, B., Burns, S., et al., 2016);
- mHealth technologies are rapidly becoming an integral part of the modern society (O'Grady, L., and colleagues, 2009);
- Health interventions efficacy and effectiveness are usually tested in RCTs. However, RCTs have a long time lag from the initiation to the outcome publication. In mHealth interventions the time lag is crucial, as these technologies may become obsolete before the experiment is completed. In many cases, the evolving nature of mHealth apps means that some components need to be continuously improved during a trial (Kumar, S., et al., 2014);

TQ2: Which are the key factors that mHealth app developers have to consider in order to properly assess their technology?

This answer can be found on the third chapter (4.3). The key factors, mHealth developers have to consider in order to properly assess their technology are:

- Ongoing evaluation throughout the app development cycle. As mobile usage becomes increasingly sophisticated and personalized, broad assessment plans are crucial in determining the efficacy of mHealth technologies (White, B., Burns, S., et al., 2016). To generate evidence, assessment needs to be conducted throughout the implementation of a mHealth app, not only at the conclusion (Lobo, R., Burns, SK., & Petrich, M., 2014);
- Emphasis on three different stages: (1) formative (while planning the app intervention), (2) summative (during the app implementation) and (3) outcome evaluation (studying the outcomes of the app intervention) (O'Grady, L., and colleagues, 2009);
- An individual's usage and engagement with a mobile app is so complex that several factors can impact the results of the intervention. To determine its success, assessment plans should consider both usage feedback and outcome measures (White, B., Burns, S., et al., 2016; O'Grady, L., and colleagues, 2009) Thus, definition of relevant outcomes, measurement techniques and criteria of assessment are crucial;
- The complex web of components that can impact the result of a mobile intervention (such as: robustness of the technology, engagement strategies, interaction of the user with the technology, and design) (O'Grady, L., and colleagues, 2009).

TQ3 (if applicable): Use the previous framework and insights to develop your own app assessment plan.

In this case, students will be able to apply the presented framework to develop and evaluate their own project. It may be important to give them access to the original paper "Collaborative Adaptive Interactive Technology" written by O'Grady, L., and colleagues (2009).

The framework that proposes formative, summative and outcome evaluation for adaptive and interactive applications is described in detail below. In order to implement it, students must follow the instructions described in the evaluation tasks. In some cases, there may not be sufficient resources to conduct an assessment that addresses all

the components listed below. In these situations, we suggest students to focus on the areas of interest – people, content, technology, computer-mediated interaction, and health systems integration – that are the most relevant to the objectives of the project/course.

Table 3 – Evaluation schema: collaborative, adaptive and interactive technology (O'Grady, L., and colleagues, 2009)

	Formative *Development and laboratory testing*	**Summative** *Efficacy and goal achievement*	**Outcome** *Impact assessment*
People	• Identification of Stakeholders' characteristics and needs *(Evaluation tasks: formal needs assessment. Identification of key characteristics of potential users, e.g. through individual interviews or the creation of personas)* • Assessment of stakeholders' interests *(Evaluation tasks: consultation with relevant stakeholders. Must assess informational needs and broader interests)*	• End users' traits *(Evaluation tasks: assessment of the end users' characteristics, i.e. computer proficiency, demographic or disease characteristics, health literacy, mHealth literacy)* * • End users' perspectives *(Evaluation tasks: collect feedback from end users regarding their intentions/motivations to use the app and their satisfaction with the intervention)* * **MARS Questionnaire*	• Patient outcomes *(Evaluation tasks: investigate the impact of the intervention on the user, i.e. psychosocial well-being, health behaviors and physiologic outcomes)* • Impact on interpersonal relationships *(Evaluation tasks: assessment of any changes/lack of changes within the patient-provider/caregiver-patient relationship)*
Content	• Quality and credibility *(Evaluation tasks: evaluations of how content accurately represents available evidence, e.g. literature review)* • Utility *(Evaluation tasks: assess completeness, relevance and*	• Quality and credibility *(Evaluation tasks: assess users' perceptions of these attributes, e.g. surveys. Do the users find the content trustworthy and believable?)* * • Subjective utility *(Evaluation tasks:*	• User-generated content *(Evaluation tasks: assessment of user-generated content on its form – narrative, numerical - and nature – advice, opinion, personal information or personal support, e.g. app analytics)*

	understandability. These parameters must be evaluated with standardized metrics and by experts or members of the target population, e.g. think-aloud walkthroughs)	*assess users' perceptions of these attributes, e.g. surveys. Overall assessment of the usefulness of the information on the app)* * • Level of personalization *(Evaluation tasks: assess users' perceptions of these attributes, e.g. surveys. Is the information applicable and usable to the end user as an individual?)* * * MARS Questionnaire	• Positioning of user-generated content *(Evaluation tasks: evaluate how the content provided by users is framed within the site. E.g. is there any mechanism for feedback/dialogue between app users and physicians/researchers?)*
Technology	• System robustness (i.e. system performance, functionality and features, privacy and security) *(Evaluation tasks: alpha/beta versions testing)*	• System reliability and speed *(Evaluation tasks: collection of usage statistics and app analytics)* • Positioning within current technology • Standards compliance	• Dynamic evolution *(Evaluation tasks: assess how well the app responds to technological or societal trends, e.g. software updates and fit to the end users' needs)* • Collaborative development models *(Evaluation tasks: evaluate alternatives for the app improvement process, e.g. open source innovation)*
Computer-Mediated Interaction	• Usability, accessibility, sociability, interactivity, and information architecture *(Evaluation tasks: alpha/beta testing,*	• User perspectives on usability and accessibility * • Demonstrated sociability and interactivity * • Collaboration * • Findability *	• Community development * • Evolution of collaboration * * *(Evaluation tasks: evaluation of summary statistics*

	heuristic evaluations and usability testing with sample populations of target users, e.g. think-aloud walkthroughs)	* *(Evaluation tasks: users' feedback, e.g. consultations with users, online surveys, measures of user activity…)*	*and longitudinal analysis of evidence of collaboration within the app that include collaborative features)*
Health Systems Integration	• Definition of evaluation metrics and process *(Evaluation tasks: whether/how well evaluation was incorporated into the design, development, and app implementation process)* • Ethics/liability *(Evaluation tasks: address legal concerns if existent)*	• Administration (services utilization, care coordination, patient safety) *(Evaluation tasks: assessment of how the app affects services utilization – usage rate for healthcare system/community services - care coordination – ways in which the app might be affecting delivery of health services - and patient safety – assessment of how/whether the app is affecting patient safety indicators such as the appropriate use of medications, e.g. questionnaires)*	• Public impact *(Evaluation tasks: assess general effects the app may have in the community)* • Cost-effectiveness *(Evaluation tasks: evaluate the existence of incremental health gains from the use of the app, e.g. economic analysis)* • Intended effect *(Evaluation tasks: suitability of the app as a mean to achieve the overall project goals)*

6 Conclusions, Limitations and Future Research

6 1 Conclusions

The case study highlights the need for health-tech companies to evaluate and properly measure their products, alongside with the acknowledged necessity to produce evidence-base understanding on the topic. Whilst there has been a significant increase on the number of trialed mHealth technologies, the majority makes use of static procedures, not suitable for complex and dynamic constructs such as mHealth app technologies. Henceforth, in order to leverage health through technology, a better understanding on how to properly evaluate efficacy, usability, behavior and health outcomes is needed.

The Weal Life desire in producing real impact, thus facilitating the caregiving experience, aligns with the need to properly evaluate these technologies.

The main research question was trying to understand why and how should health-tech companies evaluate their mHealth app technologies, as well as the key factors and methodologies these companies have to adopt in order to undertake impact measurement.

Regarding our first research question – *Synthetize why health-tech companies should evaluate their mHealth app technologies* - the study revealed that the quick proliferation and dissemination of mHealth apps across the patient spectrum outpaces the ability to properly evaluate their efficacy, security, accuracy and cost-effectiveness. Indeed, health-specific collaborative, adaptive, and interactive technologies are emerging, promising to transform roles, rights and responsibilities of all stakeholders within the system (Deshpande, A. & Jadad, A, 2006). Being the case, as any other set of interventions, mHealth technologies also carry the risk of causing more harm than good. Although these technologies may seem appealing, the lack of evidence in improving health outcomes and/or change health behavior is a barrier to the trustful and widespread adoption in the healthcare sector. Ultimately, there is no sufficient evidence-base knowledge for healthcare providers and promoters to safely incorporate mHealth apps into their plans, prescriptions or recommendations. We should assume the responsibility to "look beyond the hype, and to dissect what works and what doesn't" (Eysenbach, G., 2009), thus conducting thoughtful and careful evaluations.

Evaluation plays a critical role in high quality design, efficient development, and effective implementation of mHealth interventions. The development of appropriate frameworks will help to ensure mobile-based health technologies live up to expectations and contribute to the improvement of health. As the need for more evidence regarding the effectiveness of mHealth apps is currently acknowledged, there are significant challenges in designing and implementing evaluation plans. Regarding our second research question – *Evaluate the importance and need to adopt a comprehensive approach to mHealth apps and to undertake impact measurement* - mobile-based health interventions characteristics justify the need for a comprehensive approach to undertake impact measurement. As evolving (i.e. subject to constant updates and user experience improvements) and complex (i.e. dependent on multiple factors, such as people, content information, technology used, etc.…) technologies inserted in a dynamic and rapidly changing environment (i.e. mHealth technologies can quickly become obsolete if they don't properly adapt to the market/customers' needs or constant technological advances), mobile-based health interventions need a complete, adaptable assessment plan covering all the development stages and the multiple factors the success of the app is dependent on. This methodology assures ongoing monitoring and improvement throughout the app development cycle, preventing the app to be launched with obsolete or irrelevant characteristics, whilst generating evidence-base regarding its efficacy, usability and cost-effectiveness. Additionally,

Ultimately we approached our third research question – *Examine how mHealth apps should be measured* - by developing a holistic, multi-staged evaluation framework (based on the framework proposed by O'Grady, L., and colleagues, 2009) to properly measure and evaluate mHealth app interventions. With an emphasis on ongoing and adaptable assessment, the methodology organizes (1) formative (i.e. while planning the app intervention), (2) summative (i.e. during the app implementation), and (3) outcome evaluation (i.e. studying the outcomes of the app implementation) into five major areas of interest, (1) people, (2) content, (3) technology, (4) computer-mediated interaction, and (5) health systems integration. Applying this useful tool should help researchers to generate comparable results and superior evidence-based knowledge. As a consequence, impact measurement not only improves accountability, transparency and contribute to

build legitimacy, but mostly, helps to oversee processes of value creation to health-tech companies.

6 1 Limitations and Future Research

As the study only starts to identify some important characteristics and procedures, expanded research methods have to be deployed in order to develop more powerful and standardized tools. Therefore, some limitations of the research should be considered. Firstly, *The Weal Life* is still evaluating their app performance. As a result, the assessment plan proposed here has not been fully tested. It is hoped that as the evidence base for mHealth interventions develops, the applicability of the assessment plan will be tested. Secondly, future research is required to determine the suitability and reliability of the plan across multiple mHealth and other app domains. Additional work might have to be undertaken to ensure the generalizability of the framework proposed. Future improvements of the evaluation plan terminology and items are likely to be required, as the functionality of mHealth apps evolves.

List of References

Agarwal, S., Lefevre, A., Lee, J., et al., 2016. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *British Medical Journal.*

Bailey, C., Belter, T., Pandit, U., Carpenter, M., et al., (2013). The availability, functionality, and quality of mobile applications supporting medication self-management. *Journal of the American Medical Informatics Association.* 21(3): 542–546

Barnhart, X., Haber, J. & Lin, I., (2007) An overview on assessing agreement with continuous measurements. *Journal Biopharmaceutical Statistics.* 17(4): 529–569

Bédard, M., Molloy, D., Squire, L., Dubois, S., et al., 2001. The Zarit Burden Interview: A new short version and screening version. *The Gerontologist.* 41(5): 652-657

Boudreaux, E., Waring, M., Hayes, R., et al., 2014. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *PubMed.* 4: 363-371

Brand, L., Krebs, P., & Duncan, D., 2015. Health app use among US mobile phone owners: A national survey. *JMIR Publications. 3(4)*

Breton, R., Fuemmeler, F. & Abroms, C., (2011). Weight loss: There is an app for that! But does it adhere to evidence-informed practices? *Translational Behavioral Medicine: Practice, Policy and Research.* 1(4): 523–529

Caregiver Statistics: Demographics (2015). [Online]. AARP Public Policy Institute. Available at: https://www.caregiver.org/caregiver-statistics-demographics [Accessed: October 29th, 2016]

Chang, Y., Ghose, S., Littman-Quinn, R., et al., (2012). Use of mobile learning by resident physicians in Botswana. *Telemedine Journal and eHealth*. 18(1): 11-15

Comstock, J., (May 29, 2014). One third of docs recommended a health app to patients. *MobiHealthNews*. *http://www.mobihealthnews.com/33594/survey-one-third-of-docs-recommended-a-health-app-to-patients*

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., et al., (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of Medical Internet Research*. 15: 247

El-Gayar, O., Timsina, P., Nawar, N. & Eid, W., (2013). Mobile applications for diabetes self-management: status and potential. *Journal of Diabetes Science and Technology*. 7: 247–262

Eysenbach G. (2009). Call for Papers: Medicine 2.0. *Journal of Medical Internet*

Free, C., Knight, R., Robertson, S., Whittaker, R., et al. (2011). Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single- blind, randomised trial. *Lancet* 378: 49–55.

Free, C., Philips, G., Galli, L., et al., 2013. The effectiveness of mobile-health tecchnology-based health behaviour change or disease management interventions for health care consumers: A Systematic Review. *PubMed*. 10(1)

Free, C., Whittaker, R., Knight, R., Abramsky, T., et al. (2009). Txt2stop: A pilot randomised controlled trial of mobile phone-based smoking cessation support. *Tob Control*. 18: 88–91

Fukuoka, Y., Komatsu, J., Suarez, L., et al., (2011). The mPED randomized controlled clinical trial: applying mobile persuasive technologies to increase physical activity in sedentary women protocol. *BMS Public Health*. 11(1): 933

Grady, L., Witteman, H., Bender, J., et al., 2009. Measuring the impact of a moving target towards a dynamic framework for evaluating collaborative adaptive interactive technologies. *Journal of Medical Internet Research.* 11: 1-10

Istepanian, R., Laxminarayan, S., & Pattichis, C., (2006). M-health (p. 619). *Springer.* Retrieved from http://link.springer.com/content/pdf/10.1007/b137697.pdf

Kratzke, C., & Cox, C., 2012. Smartphone technology and apps: rapidly changing health promotion. *Global Journal of Health Education and Promotion.* 15(1)

Kumar, L., Sharanie, K. & Jaspajleet, D., 2015. Barriers to adoption of consumer health informatics applications for health self-management. *Health Science Journal.* 9: 1-7

Kumar, S., Nilsen, W., Abernethy, A., et al., 2014. Mobile Health Technology Evaluation. *American Journal of Preventive Medicine.* 45(2): 228-236

Laxminarayan, S., & Istepanian, S., (2000). UNWIRED E-MED: the next generation of wireless and internet telemedicine systems. *IEEE Trans. Inf. Technol. Biomed.* 4(3): pp. 189–193

Lester, T., Ritvo, P., Mills, J., Kariri, A., et al. (2010). Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. *Lancet* 376: 1838–1845

Lobo, R., Burns, K. & Petrich, M., (2014). Supporting health promotion practitioners to undertake evaluation for program development. *BMC Public Health.* 14: 1315

Martínez-Pérez, B., de la Torre-Díez, I., Candelas-Plasencia, S., & López-Coronado, M. (2013). Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. *Journal of Medical Systems.* 37(5): 9976

McCubbin, H.I., Patterson, J., & Glynn, T. (1982). Social Support Index (SSI). In H.I McCubbin, A.I. Thompson, & M.A. McCubbin (Eds., 1996), *Family assessment: Resiliency, coping and adaptation. Inventories for research and practice* (pp. 357- 389). Madison, WI: University of Wisconsin System.

mHealth Economics Research Program (2016). mHealth App Developer Economics 2016: The current status and trends of the mHealth app market. *Reseacrh2guidance 2016.* http://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2016.pdf

Mohr, D., Schueller, S., Riley, W., et al., 2015. Trials of intervention principles: Evaluation methods for evolving behavioral intervention technologies. *Journal of Medical Internet Research.* 17

Muessig, E., Pike, C., Legrand, S., & Hightow-Weidman, B., (2013). Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. *Journal of Medical Internet Research*. 15: 1

Pagoto, S., Schneider, K., Jojic, M., DeBiasse, M., et al., (2013). Evidence-based strategies in weight-loss mobile apps. *American Journal of Preventive Medicine*. 45: 576–582

Silva, B., Rodrigues, J., et al., 2015. Mobile-health: A review of the current state in 2015. *Journal of Biomedical Informatics.* 56: 265-272

Stoyanov, S., & Psych, M., 2015. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR Publications.* 3(1)

U.D. of Health, H.S. Food and Drug Administration, Radiological Health, (2015) Mobile medical applications: Guidance for industry and food and drug administration staff http://www.fda.gov/downloads/MedicalDevices/../UCM263366.pdf

Vervloet, M., vanDijk, L., Santen-Reestman, J., et al., (2011). Improving medication adherence in diabetes type 2 patients through Real Time Medication Monitoring: A randomized controlled trial to evaluate the effect of monitoring patients' medication use combined with short message service (SMS) reminders. *BioMed Center Health Services Research.* 11(5)

White, B., Burns, S., Giglia, R., et al., 2016. Designing evaluation plans for health promotion mHealth interventions: a case study of the Milk Man mobile app. *Health Promotion Journal of Australia. 27:* 198-203

WK Kellogg Foundation, (2004). *WK Kellogg Foundation evaluation handbook.* WK Kellogg Foundation

Wolf, A., Moreau, F., Akilov, O., et al., (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatol.* 149: 422–426

World Health Organization Global Observatory for e-Health: mHealth: New horizons for health through mobile technologies: second global survey on eHealth, (2011). http://www.who.int/goe/publications/goe_mhealth_web.pdf

Zarit H., Orr, K., Zarit, M., 1985. The hidden victims of Alzheimer's disease: families under stress. *New York: New York University Press*