10-2013

Theoretical, Methodological and Practical Challenges in Designing Formative Evaluations of Personal eHealth Tools

Michael S. Dohan
Lakehead University, msdohan@lakeheadu.ca

Joseph Tan
McMaster University, tanjosep@mcmaster.ca

Theoretical, Methodological and Practical Challenges in Designing Formative Evaluations of Personal eHealth Tools

Michael S. Dohan
Faculty of Business Administration
Lakehead University
msdohan@lakeheadu.ca

Joseph Tan
DeGroote School of Business
McMaster University
tanjosep@mcmaster.ca

Abstract: This article describes the challenges experienced in designing a formative evaluation of a mobile personal eHealth tool for disease symptom tracking, in the context of Inflammatory Bowel Diseases management. As this unique context is characterized by regular visits with physicians, and the need to report disease activity to the physician, a mixed methodology is used to assess the fit of the technology to the context, and a multilevel model is used to evaluate theoretical links between patient and physician. Other issues include selection of sampling techniques, recruitment, and preventing study attrition by participants.

INTRODUCTION

Advances in technology are increasingly applied to healthcare processes, in particular, those that enhance and empower the care received by patients (Eysenbach, 2001). Personal eHealth tools (PETs), such as personal health records (Archer, Fevrier-Thomas, Lokker, McKibbon, & Straus, 2011), entail the delivery of healthcare services and resources via technology directly to patients. The capabilities afforded to the patients through devices that are affordable, available, and versatile, and have provided health benefits in the context of diabetes self-management (Glasgow, 2011), and aided in maintaining general personal medical documentation (Tobacman et al., 2004). Further, the PETs can be applied in a way that can enhance the connection between patient and provider, through capabilities such as communication of symptom data (Johansen, Henriksen, & Berntsen, 2011; Johansen, Henriksen, Horsch, Schuster, & Berntsen, 2012).

As the benefits to using these technologies needs to be clearly documented, they require careful evaluation. The Randomized Controlled Trial (RCT) experimental design is regarded as the "gold standard" for the evaluating the efficacy of medical interventions (Kaplan, 2001), essential for adoption of the technology. However, the RCT and other experimental designs may not be sufficient for evaluating PETs, in particular before a widespread implementation takes place, for several reasons (Easterby-Smith, Golden-Biddle, & Locke, 2007; Kaplan, 2001). First of all, experiments are a costly and complicated undertaking. Second, experiments are better suited for interventions that are simple in nature. For instance, a drug and a mobile application differ greatly in how it will be used by the user. A PET application requires more learning to use in an effective manner, and is usually contingent on many more factors than those purely biological. Third, as the purpose of an experiment is to test the viability of an intervention to produce a significant outcome (Cook, Campbell, & Peracchio, 1990), it often ignores the context of use due to a high focus on quantitative methods, which is important in answering questions as to how the intervention is used, what its barriers to use are, and what features are important. These reasons compel researchers to conduct formative evaluations (Campbell, 2000; Craig et al., 2008), which tend to focus on informing the development process, and are thus conducted earlier on in development, informing subsequent experimental studies.

In order to conduct formative evaluations on complex interventions such as PETs, it is important for researchers to consider the context in which it is used, in order to generate knowledge and build theory for PETs, despite the impracticalities with using experimental design. One way to further the wider field of technology in healthcare applies here (Kaplan, 2001); research that asks the question of *how* a particular technology is used in a certain context can be used to build knowledge as to how the technology relates to or impacts the existing care processes established by the patient. Specifically, it can ask the question as to *what factors* are important in the context, be they psychological, behavioural, social, cultural, relational, related to workflow, institutional influences, or other

pertinent factors. Further, it can help us understand *how* an intervention is implemented (Oakley, Strange, Bonell, Allen, & Stephenson, 2006). Researching information systems in such a way is referred to as the "fit" (Easterby-Smith et al., 2007; Kaplan, 2001) that an information system has with these and other factors in the environment. It is argued that researching the fit of technology with its environment is the direction best suited to generate knowledge in this area, but the question remains as to how the theory and protocol is configured. Given the multitude of unique environments in healthcare that can benefit from technology, guidance is valuable for researchers in this area.

The purpose of this paper is to present theoretical, methodological and practical challenges experienced when designing formative evaluations to assess the fit of patient-centric eHealth tools to their context of implementation. Drawing on several established approaches, it is the goal of this article to provide guidance to researchers in constructing theoretical frameworks for their research, as well as designing a suitable methodology, so that the product of their endeavours is scientifically rigorous, valuable to practitioners, and efficiently executed. To this end, this paper will first describe PETs as a complex intervention, the degree to which they must fit their context, as well as how they can be evaluated to improve fit. Following, challenges in determining design for an evaluation will be described, including selection and integration of theoretical frameworks, choice of methodology, and issues associated with selecting and managing participants. The experiences designing an evaluation of a PET, specifically to record and report gastrointestinal related symptoms to physicians, within the context of Inflammatory Bowel Diseases (IBD) care, will be used to illustrate how these challenges can be addressed when designing a study protocol.

BACKGROUND

PETs as Complex Interventions

It is difficult to provide an exact definition of a complex intervention, although guidelines are used for their identification (Campbell, 2000; Craig et al., 2008). Complex interventions are deemed so as they contain several interacting components (Campbell, 2000; Craig et al., 2008). For further elaboration, they provide several characteristics of a complex intervention. These include, but are not limited to, number of components of the intervention, interactions between the user and intervention, groups targeted by the intervention, number of outcomes, and degree of flexibility of the intervention (Campbell, 2000; Craig et al., 2008). Further, it may be difficult to identify exactly what components contribute to the outcomes of the use of a complex intervention (Campbell, 2000; Craig et al., 2008), yet containing the components into one intervention would have a synergistic effect where the sum of the components is less than the entire intervention (Casti, 1997). A "stroke unit" is provided as an example of a complex intervention (Campbell, 2000). A stroke unit may provide several services, contain many workers, and produce several positive outcomes, yet it is hard to define what the actual components of the stroke unit are, which are beneficial, or even if the benefit would still exist if the components were provided elsewhere.

Personal e-health technologies (PETs) is defined broadly for this article as technologies targeted to patients, healthcare consumers, or the population at large, for a health related purpose. This includes electronic personal health records (ePHRs, Archer et al., 2011) or any other intervention used for health-related data collection and management that predominantly involves the patient. As many of these interventions are comprised of several components, serve several purposes, and can be implemented on various technologies, such as the web, or mobile phones of various platforms, they can be considered complex interventions. Further, some interventions may involve integrating users with other healthcare professionals or healthcare system processes. One case is that of Electronic Symptom Reporting (ESR, Johansen et al., 2011, 2012), where patient technology is used to collect data on disease activity, which is subsequently transported to a physician or other healthcare practitioner. Mobile devices themselves afford benefits to users that are not necessarily tied to any software feature, and can be considered a component of an intervention in themselves. Generally speaking, a large amount of PETs will certainly be considered complex interventions, attaching with them implications of their evaluation.

Fit

Due to the issues with complex interventions, a different and varying approach must be taken in their evaluation. Unlike simpler interventions, such as a medicine in pill form, experimental designs such as randomized controlled trials (RCTs) do not account for much of the contextual information associated with the intervention. This can include existing processes, approach to implementation, socio-cultural and other environmental factors that may impact uptake of the intervention. This can be illustrated by comparing a pill to a PET implemented on a mobile phone, for the purpose of helping a patient manage their chronic disease. The former typically requires easily understandable, clearly defined instructions and outcomes for the user ("take twice a day with food", for example). The latter may require the user to learn how to use mobile devices, learn how to enter, interpret and act upon symptom data and reports, respond to alerts, enter personal information, and integrate or synchronize with other systems, to various benefits to the user and other people. Given the relative complexity of the latter intervention, more knowledge may be necessary in its implementation that will be used to ensure adoption, long term use, and long lasting benefits. This knowledge will typically be used to discover the requirements for the intervention, barriers to implementation and to guide further development of the technology in question (Craig et al., 2012). Use of a purely experimental research design typically is employed to demonstrate that use of intervention has a significant positive impact on the sample, and that the benefit is generalizable to a defined population (Woolard, 2004). Although it is important that any medical or healthcare intervention demonstrate such a benefit, a purely experimental methodology is poorly equipped to address these important aforementioned issues.

Health technology researchers have addressed the difficulties of evaluating complex interventions through focusing on the "fit" of technology to its context. As termed by Kaplan (2001), the "fit" refers to the compatibility of a technology with the context that it is implemented in, and there are many possible aspects of the context which may be relevant when determining fit. For instance, the technology may have to fit with workflow (Yusof et al., 2008; Ammenwerth, Iller & Mahler, 2006), or culture (Kaplan, 2000), for example. Indeed, a substantial amount of frameworks have been developed to attempt to explain the fit of systems in their respective environments. These studies often take a sociotechnical (ST) theory (Edwards, 1972) approach. ST theory has focused on the fit of systems to their context has been imperative in the success of systems in various contexts. Each respective context may require the system to fit the context on a unique set of dimensions, as reflected in each of the ST models. For instance, Sittig and Singh (2010) propose that healthcare information technology must fit the context on the following dimensions – computing infrastructure, clinical content, system interface, relevant people and roles, workflow and communication, internal policy and culture, external rules, and performance measurement. As systems typically implement their own processes, procedures and other various relevant factors, they must be in congruence to the context on all of these dimensions, or there is a risk that the users will not adopt the system, resulting in wasted efforts on building the system. In other words, systems are not guaranteed to meet the goals they intend to; it is entirely possible that they have no effect, or worse yet, exacerbate the problems they intend to address. Care must be made in ensuring that the system will be effectively used, and contribute to positive outcomes that it intends to address.

Formative Evaluation of Complex Interventions

As elaborated on before, experimental designs on their own are insufficient to determine the fit that a system has with its context, rather experimental research is largely used to establish that an intervention has had a significant benefit on an outcome through quantitative methods. Following a call for a variety of methodologies to be used in evaluating health technologies in order to ensure sufficient knowledge of the context (Kaplan, 2001), qualitative methods (Lee, 1999; Bachiochi & Weiner, 2004) can be employed for this purpose, and mixed methods research (Johnson, Onwuegbuzie & Turner, 2007) can be used to combine performance measurement and contextual description into one study. However, the use of qualitative methods alone cannot ensure that an intervention will be successful. There must be a mechanism to inform those responsible for crafting the intervention on the shortfall between the context of application and the qualities of the intervention, so that a change may be implemented in the intervention. Having such will mitigate the risk of an intervention falling into disuse.

Determining fit of a healthcare technology intervention with the intended context can be performed with a formative evaluation (Campbell, 2000; Craig et al., 2008). As formative evaluations are largely used to inform developers of an intervention, they typically take place early on in the implementation process, before a larger implementation or evaluation takes place. The developers then can act upon the results of the formative evaluation by improving the

artifact before it is implemented. In the case of technology, they may find out that the inclusion of a certain feature is a crucial requirement of the user, or what outcomes the users expect from the use of the technology (Flagg, 2013; Scriven, 1967; McClintock, 1984). A formative evaluation can uncover this requirement before a large scale rollout, and the developers can choose to improve the application (Brown & Kiernan, 2001). A formative evaluation is often conducted under a larger implementation framework, where subsequent summative evaluations will report on performance, such as cost, and outcomes, such as relevant health outcomes (Patton, 2008).

As formative evaluations are regarded as an important activity in crafting successful complex interventions (Craig et al., 2012), it is important that they are conducted. Although these types of evaluations are typically smaller scale than summative interventions, they still present several unique conceptual, methodological, and practical challenges that must be addressed by the researcher. The remainder of this article will discuss some of these issues for consideration by researchers involved in formative evaluation of PETs in a healthcare context. An illustration will follow on how these issues can be addressed in a formative evaluation of a smart phone application for the purpose of Electronic Symptom Reporting (ESR) of a chronic disease patient symptom data to a healthcare practitioner.

CONCEPTUAL CHALLENGES

Developing a Theoretical Model

A theory is an organization of knowledge, an attempt to describe a defined context in the real world, bet it based on a behavior, situation, or outcome of a process (Dubin, 1976). A theoretical model consists of two components, at minimum. First, it specifies the units important to the theory as variables, as well as descriptions of relationships among the various variables. Second, it describes the context in which the theory is relevant to, for which it is proposed that this theory will be supported. If the theory is to be empirically tested, an empirical indicator is assigned to each of the variables, which, when data has been collected, can be used to test the relationships between variables, referred to as hypotheses. A theory can be used to predict events, explain why events occur, gain a deeper understanding of a context, or to describe an ideal situation (Dubin, 1976; Gregor, 2006). There are several qualities that good theories possess. First, as it is impossible to explain every phenomena in every situation, it must be falsifiable, so it can be tested and possibly refuted (Popper, 1963). Stronger theories are tested, with the purpose of falsifying it, but each time the author fails to falsify the theory it adds strength to the theory. Good theories are supported by existing empirical evidence. Second, a theory must be parsimonious, in order that can be easily and simply communicated and understood (Cohen, 1990). As well, good theories are appealing and make sense to the practitioner (Dubin, 1976). A theoretical model should be useful when applied to the real world, and fulfilling to practitioners in guiding the endeavors (Dubin, 1976; Gregor, 2006). It must be parsimonious, and focused. As well, it must be generalizable, or useful in other situations to explain, predict, or gain understanding there (Gioia & Pietre, 1994).

Healthcare is very rich and diverse in context, therefore there are many opportunities and efforts to explain behavior and develop theory. There are many specialties, many diseases, many healthcare system functions, disciplines and professions, payers, research traditions, and cultures represented in healthcare, making for a diverse system. It can explain clinical and organizational behavior, select or tailor interventions given a certain problem or context, evaluation of implementations. In public systems, the taxpayer is relevant, whereas shareholders and other stakeholders are relevant in private systems. Theory in healthcare hasn't been widely employed as they have in other areas (Walshe, 2007). Hence, there has been an influx of theories developed outside of healthcare, such as complex adaptive systems theory (McDaniel, Lanham & Anderson, 2009), actor-network theory (Greenhalgh & Stones, 2010) and technology adoption theory (Gallant & Boone, 2009). Indeed theories are important in healthcare, yet their development poses a challenge. The evidence based medicine approach espouses the importance of theory development if innovations and research are to be implemented in healthcare systems (Walshe & Rundall, 2001; Guyatt et al., 1992). One way of ensuring that theory is used in healthcare is to ensure that it contains a practical or strategic element. Also, the link between theory and practice must be clear. That way, it is more likely that it will eventually be applied. This can be accomplished by including both manager and health researchers in the theory development process, in particular during initial phases of research.

The researcher may see it necessary to integrate several theories in the interest of seeking a higher understanding of the context, thereby creating a unique theory for use in a highly unique context. Further, when applying theory from outside of healthcare, the research may find that there is little empirical evidence to support its use within healthcare.

Even in an early formative assessment, preferably during development, it is important to adhere to a theoretical perspective (Craig et al., 2012), but a suitable theory, or sufficient empirical support, may not exist. In the case that researchers are attempting to utilize a theory where none exists before in this context, it is advisable to start with an exploratory phase in order to determine the important factors and outcomes to the context being studied, and to start on a smaller scale study, to start with preliminary support for the theoretical model (Power et al., 2004; Rudolf et al., 2006)

Multilevel models can be used where there is a relevant hierarchical component to a context (Hoffman, 2004; Dixon & Cunningham, 2009). For example, a physician can have many patients, but a patient can only have one physician. This entails that the relationship between patient and physician may have an impact on certain outcomes of care, such as patient satisfaction (Barr et al, 2003). Indeed, a tenant of patient centered care (Wagner et al., 2005) is to foster a relationship with the patient so that the physician gains a higher understanding of the personal situation, preferences and knowledge held by the patient, for the purpose of enabling the patient to participate as a decision maker in their own care decisions. In some contexts, especially those where long-term chronic disease management is relevant, the use of PETs may enhance or otherwise augment the relationship between patient and provider. For instance, an attribute of the group of patients of a certain physician can impact the individual physician in some way (Dixon & Cunningham, 2009), or the individual physician can impact the group of patients (Hoffman, 2004). Other hierarchical structures can include physicians within a department, hospitals within a region, or medical students within a school. As multilevel theories have been applied successfully in understanding healthcare contexts (Piatt et al., 2006), the researcher must consider the contribution of using them when evaluating healthcare technology (Kaplan & Harris-Salamone, 2009).

METHODOLOGICAL CHALLENGES

Selection of Methodology and Design

In formative assessments, the target outcome of the intervention is not always necessarily known. In complex interventions, this challenge is exacerbated, as different stakeholders of the intervention may have their own goals, or a user may have two outcomes in mind. For example, if a single complex intervention is implemented for both a physician and a patient, the patient may want to increase their ability to cope with a chronic disease, whereas the physician may want to increase efficiency in the delivery of care to a population of patients. A purely quantitative evaluation of this intervention will measure only a limited number of outcomes for economic reasons, but other outcomes may be appropriate to the user, context, or system, and would not be measured or otherwise detected. The use of quantitative methods (Creswell, 2009) function to construct a rich description of the perspectives, attitudes, experiences, and goals of the users of the complex intervention, otherwise lost to the purely quantitative study. Results of these types of studies serve to inform future summative evaluations, and develop theory in these particular areas.

Mixed methodologies offers an enhancement over pure qualitative studies (Creswell, 2009). Two general categories of pragmatic research exist: sequential, with each part executed as a phase, and concurrent, with both parts being executed at the same time (Creswell, 2009; Easterby-Smith et al., 2012). The sequential can take one of three forms: explanatory, exploratory, and transformative. In the sequential explanatory form, the results of a quantitative phase are enhanced by qualitative data from participants in the first. This configuration is valuable for enhancing unexpected results (Morse, 1991), or for further interpretation by outliers in the first phase. In the sequential exploratory configuration, the results of a qualitative phase are followed by a quantitative phase, which is quite useful when a researcher wishes to develop a concept and an instrument in the same project. The sequential transformative strategy employs a critical theory lens to guide the project, and may employ quantitative or qualitative in any order. In any of the sequential strategies, an a priori theory may or may not be employed, and dominance is usually given to the first phase. The concurrent strategy can take the form of triangulation, embedded, or transformative. Concurrent triangulation strategy entails the integration of quantitative and qualitative data after both datasets are collected. Concurrent embedded strategy prioritizes one result set over another, which uses the secondary result set to enhance the results from the primary. Concurrent transformative strategy uses any of the configurations above, with the researcher adhering to a critical theory lens to guide the research.

The value of conducting mixed methods research exists in the fact that results are enhanced to a level greater than any of the individual components on their own (Farquhar, 2011). For instance, if an outcome of use of a PET is

found to be lower than expected, an exploratory phase in the same study can be used to determine why this is, what barriers exist to effective use of the PET. As mixed methods research is significantly more complex than purely quantitative or qualitative, justifying mixed methods research projects is important, despite the clear benefits of these methodologies. These research projects can consume more resources, take longer to complete, and tax participants to a higher degree. Further, they may not readily be understood by reviewers of conferences, journals, and granting agencies, restricting opportunities for the researcher. With this in mind, the reason to use mixed methods must be clearly demonstrated.

Recruitment of Participants

The acquisition of potential participants in formative evaluations presents challenges for the researcher, although the difficulties with recruitment associated with experimental designs, such as randomization or blinding (eg. Shcherbatykh, 2008), are less of a concern for smaller scale, non-experimental evaluations. Although PETs can be integrated with the system of the healthcare provider to several degrees (Tang et al., 2006), this discussion will be limited to integrated PETs, as patient-physician interaction is much less relevant with standalone PETs (ie PETs that in no way interact with a healthcare provider's system). Tethered PETs would allow healthcare providers to grant read-only access to certain data to the patient, whereas interconnected PETs allow for richer data interchange between patient and provider (Tang et al., 2006).

An example of an integrated personal healthcare tool is DiaMonD (Wickramasinghe, Troshani, & Goldberg, 2010). DiaMonD is a smart phone application for the purpose of sending symptom data of patients with diabetes directly to the healthcare provider. In this sense, the tool is an intervention that involves several people (patient and physician), each with their own activities (patient must enter symptom data, and physician must receive it) and each user can perform several main functions. In this case, the patient can either enter or view their own data, and the physician can view the data of one person, or aggregated data of their patient list. When developing this complex intervention, it is important to involve all people in the care process. In this case, it will be both the diabetic patient, and their physician. It is clear to see that, since DiaMonD can potentially benefit both diabetic patient and their doctor, that both could be targeted for research, whether in the same project or not.

For studies that are not intended to gain insight into a larger population, the use of non-probability sampling methods, such as convenience, self-selection, or snowball sampling, can be acceptable (Small, 2009). If access is granted, patient registries has been identified as a valuable sampling frame for recruiting patient participants in healthcare studies (Snyder et al., 2009) although these are not available for every patient group. Other outlets may be of value, such as patient associations, newsletters, websites, social media groups, local associations and foundations. Of course, physicians may allow access to their patient population, in particular if a physician is a partner in the research project. Also, out-of-date registries pose a problem. If patients do not suffer from the disease anymore, they are unlikely to keep in touch with a patient registry (Patel, Doku & Tennakoon, 2003). If the researcher is studying a condition with a high death rate, the registry may be full of records of deceased people. As well, if the budget allotted for the study is low or non-existent, the researchers may require the patients to own their own technology, which may reduce the number of potential participants.

Apart from patients, there may be a need to recruit a significant number of physicians, which presents some difficulties (Johnston et al., 2010). First, a sampling frame for physicians of a certain specialty may not be kept accurate and up to date, potentially resulting in unplanned time and effort spent assessing participants for eligibility, or changing the definition of eligible participants. Second, clinic staff may have to be engaged to perform some duties associated with the study, such as recording participant data, or advertising for the study. Facilitators to recruiting physicians to studies included to have a rapport with the clinic staff, employ clinicians to recruit other clinicians, as well as various incentives. Incentives may include a payment of a very high sum of money, continuing medical education credits, and provision of other clinical resources. As well, the physician participant must perceive that the study is important and relevant to their practice, in order to achieve buy-in. Personal connections between the researchers and physician participants is regarded as important as well.

Retention of Participants and Addressing Attrition

In research, attrition refers to the reduction of the number of participants in a study. When conducting a longitudinal study, researchers will typically require the participants to use an intervention for a period of time, and submit data

after the time period has elapsed. When the time period has elapsed, it is probable that a fraction of the participants that started the intervention will not perform the activities at the end. The attrition rate then refers to the number of people that have not completed the study, divided by the total number of participants at the start of the study. Attrition hurts the generalizability of a study, for several reasons. First, if the number of participants is reduced to a level that is below the minimum sample size dictated by a power calculation (Dupont & Plummer, 1998), the ability to report that a study made inferences on a population is compromised. If the number of participants sinks low enough, the ability to perform certain statistical analysis techniques is removed. For instance, both Structural Equation Modelling (SEM) (Klem, 2000) and regression (Licht, 1995) have a minimum sample size specified in order for any analysis on the data to be considered valid. From a practical standpoint, the time, effort and resources in recruiting and signing on participants to a project is wasted when a participant quits the study (Cotter et al., 2005). For these reasons, it is in the researchers' best interests to try to retain their participants through until the end of the study.

Several approaches exist to help ensure that the attrition rate is as low as possible. First and foremost, the research design certainly can impact the attrition level of a study, such as length of an instrument (Hoerger, 2010). Beyond that, retaining participants in the study can be aided by making regular contact with the participant, scheduling callbacks, establishing a good rapport with the participant (MacLachlan, 1988; Nagler et al., 2013), and "cultivat(ing) subject loyalty" (Probstfield, 1986) in the participant, to the study and researcher. It may also help to schedule contact as part of routine healthcare, such as regular appointments (Meinert, 2012). When data must be collected, not limiting the number of times that researchers can contact participants to encourage them to complete the study will improve results (Woolard et al., 2004; Cotter et al., 2005). Limiting the times that researchers are allowed to contact the participants may negatively impact post-intervention contact with participants. In very long studies, as participants move and fall out of contact, researchers may need to employ strategies in finding lost participants, such as web searching, and multiple contact methods (email and telephone, etc.). In doing so, contact is made with the participant in order to determine any reasons for withdrawal or barriers to participation. It has also been suggested that, in conducting clinical trials, that a dropout participant can be "re-recruited" back into the study (NHLBI, 1998). Dropout participants may be valuable source of qualitative data for this same purpose, although the need to re-enroll the participant may not exist. Other approaches include collecting much of the data up front (MacLachlan, 1988).

Studying the nature of research retention is the focus of some research projects. In a study of an online weight loss intervention, it was found that younger participants and participants with higher levels of self-efficacy for weight loss would remain with the 12-month long study (Glasgow et al., 2007). In a similar study, engagement of the online intervention, measured by several website usage variables, was found to be a predictor of study retention after 3, 6 and 12 months (Couper et al., 2010). The condition of the participant studies may also affect attrition. It has been found that people with the ADD condition were easier to retain than those with CD and OCD (Cotter et al., 2005). Other factors such as mistrust of researchers, community involvement, incentives, and cultural adaptation of studies have been found to impact study retention, in particular with minority populations (Yancey, Ortega & Kumanyika, 2006). Another approach is to look at determinants of attrition that are variables related to the person, such as their beliefs and attitudes (Wojtowicz, Day & McGrath, 2013). In implementing a depression, anxiety and stress module, student participants who were more likely to adhere to the study the longest had higher perceived behavioral control in completing the prescribed tasks, use phone instead of email for their chosen method of communication, and were older in age. Further, people who dropout may be experiencing a negative outcome from use of the intervention (Ladouceur et al., 2001), resulting in a systematic bias in the results.

PRACTICAL CHALLENGES

Creating Value for Developers

Developers of technologies rely on interaction with the users to ensure acceptance of the technology by its users (Ries, 2011). In fact, close interactions are an integral part of newer software development methodologies such as Scrum (Rubin, 2012) and Agile (Beck et al, 2001). One method of communicating with users is via prototyping, where a partially functional or representative version of the system is showcased to a user, or the user is asked to use the system, with the intent of eliciting judgments on the system to inform its further development, including

features, usability, function and congruence to workflow. Formative evaluations and prototyping are similar in purpose, yet differ in aspects such as scale, stage, use of theory and formality.

A more focused explanation of how developers require feedback is described with the concept of validated learning (Ries, 2011; Rubin, 2012). Features in software are built on an assumption held by the developer that the user will find it useful. The assumption turns out to be false when it is discovered that the user will not find this feature useful. If much time has been spent building a feature that is not considered valuable by the user, the time and effort that is used to build this feature is considered waste, as it does not contribute to the user's goals in using the software. The longer that the development team uses unvalidated assumptions to inform their development, the larger the risk that efforts will be considered waste if the user does not adopt these features. This reasoning underlines the importance of validating these assumptions as early as possible, in order to ensure that development efforts are directed towards features that are more likely to be adopted by the user. Formative evaluations, pilot studies and other early evaluations should seek to validate assumptions held by the development team, in order to direct development. In order to do this, researchers may want input from developers when designing formative evaluations.

Conveying Value of the Study to Reviewers

Although the formative evaluations will primarily serve to improve the complex intention by informing its design, there are important reasons why researchers will want to communicate the research protocol and results to parties external to the research and development processes. First, the research team may want to apply for funding from a foundation, industry or government organization, such as the Agency for Healthcare Research & Quality (AHRQ) or the Canadian Institutes of Health Research (CIHR). Second, the research team may want to communicate the results of their research, including any valuable theoretical contributions, to various conferences and journals. As research in the area of eHealth is relevant to several fields, including medicine, healthcare, management, information systems, and engineering, the writing of proposals and articles must be tailored to their respective audience. The challenge arises when reviewers from one area are not familiar with languages, concepts and approaches from another. This challenge is inevitable, given the diverse number of researchers and areas contributing to eHealth research.

An example of this challenge will exist when those from the medical field review material for a formative evaluation. Problems may manifest if the reviewer is not aware that an evaluation is a formative evaluation, or the purpose of such an evaluation is largely to provide feedback for refinement of the intervention, or the value of these reviews to the developers and future researchers of an intervention. In this case, the reviewer may assume that the evaluation is indeed an experimental design, due to the popularity of RCTs in medicine, and conduct the review as such. As discussed, a formative evaluation is not suitable for demonstrating a significant positive outcome due to the use of an intervention, nor is it suitable to do so. This is accentuated when evaluating complex interventions, as a formative evaluation is essentially a precondition for experimental research. To address this challenge, the researchers must clearly communicate the definition of a formative evaluation, its purpose, its relation to other experimental evaluations, and its value, so that the reviewer does not assume that the evaluation is experimental. As well, reviewers may attempt to convince researchers to stick to more established theories. There is value in looking at a context in a new perspective, yet a reviewer may believe otherwise. For this reason, rationale for the use of this theory must be clearly provided, and communicated in such a way that can be understood by researchers from all applicable areas.

DISCUSSION – FORMATIVE EVALUATION OF "THE CDHF GI BODYGUARD"

The research proposal in question is for a formative evaluation of The Canadian Digestive Health Foundation's (CDHF) Gi BodyGuard ("Gi BodyGuard"), a smartphone application used for recording and tracking symptoms related to digestive health, in the context of Inflammatory Bowel Diseases (IBD) care, a group of chronic diseases that manifest as inflammation or ulcers within the gastrointestinal tract (CCFC, 2008; Carter, Lobo & Travis, 2004). In the absence of a definitive cause and cure, IBD patients are compelled to cope with their chronic disease in the long term (Carver, 1997), and treatment is focused on achieving remission and improving the quality of life for its patients. Gi BodyGuard is a smartphone application that is used by users to record, view and share data clinically relevant to IBD care - namely stool qualities, episodes and severity of pain, diet, medication and water intake (Robinson, 2001; Kennedy et al., 2004; Lakatos, 2009) – which can be shared with a physician during a regularly

scheduled appointment. The physician can then use the data to assess disease activity over a historical time period, and draw inferences on the efficacy of the treatment prescribed to the patient. In this sense, this can be considered an application of Electronic Symptom Reporting (ESR: Johansen et al., 2012; Dohan & Tan, 2014), where electronic means are used to report patient symptom data to physicians, so that they may gain some knowledge about patient experiences. The application is currently available on iTunes and Google Play as a free download.

Developing a Theoretical Model

A theoretical model should reflect the context of its application. In this case, the context is IBD care. Among IBD patients, the clinical features of Crohn's Disease (CD) and Ulcerative Colitis (UC) include the defecation of blood, mucous and pus, as well as the occasional sudden need to empty the bowels (Baumgart & Carding, 2007; Wilson & Greco, 2012) Whereas CD can manifest anywhere on the digestive tract, including mouth, esophagus, stomach and small intestine, UC is isolated to the rectum and the entire colon or portions thereof (Lennard-Jones, 1989). Although UC's symptoms are more homogenous than those of CD, symptoms of UC may include abdominal pain, diarrhea and weight loss (Carter, Lobo & Travis, 2004). During the IBD patient lifetime, disease activity is intermittent, occurring at different times and with varied levels of severity (Munkholm, 1995). Hence, many patients seek ways in which they can better cope with their disease (Casati et al., 2000). Without a cure, the goal of IBD treatment is to reach remission and improve patient quality of life. Treatment generally involves monitoring disease activity, and accepting various drug therapies, laxatives and topical treatments (Robinson et al., 2001). Adherence to this treatment, as well as dietary adjustment, periodic physician office attendance and active management of symptoms remain an important part of coping with IBD patients to attain remission and improve their quality of life (Robinson et al., 2001; Kennedy et al., 2004; Lakatos, 2009).

As no known theoretical framework specifically addresses ESR in the context of chronic diseases, this presented several challenges to the researchers. First, IBD patients who are likely to adopt an application for recording, viewing and sharing data relevant to coping with IBD disease activity likely have some knowledge of the disease before they start using the application. This application will have to offer some sort of benefit over and above their established disease coping processes. Second, as the chronic nature of the disease entails periodic visits with a clinician spanning a long period of time, their perspective on patient use of the application will likely influence whether the patient uses the application or not. Third, as the impact of self-management on disease activity of IBD is not yet known (Barlow et al., 2010), it can't be certain that use of the application will impact disease activity, but there still may be some other benefit to use. Lastly, the perspective held by the patient on the specific technology, as well as their general ability to learn how to use technology, will likely influence its use. In addition to these challenges, and although the Gi BodyGuard technology has underwent some formative evaluation in its development (Mulvale, 2013), it has not been particularly evaluated in the described clinical context. For these reasons, the purpose of this research is to better understand how the use of particular smartphone application is used to assist coping with disease activity within the normal process of IBD care.

The theoretical background of this formative evaluation is constructed by integrating three relevant theoretical perspectives, namely Self-Regulation Theory (SRT) (Diefenbach & Leventhal, 1996), Information System Continuance (ISC) (Bhattacherjee, 2001), and Information Quality and Satisfaction (IQS) (DeLone & McLean, 2003; Wixom & Todd, 2005; Nelson, Todd & Wixom, 2005). SRT offers an approach to understanding how people with chronic diseases process environmental and somatic stimuli in relation to their condition to learn how to cope with their long-term illness. As the need to study the fit of technology into a context is clearly articulated earlier, the authors argue that the context for the patient experiencing the disease is one in which the appraise disease related stimuli and environmental factors in relation to their disease, and how it can be better managed. ISC provides a framework for understanding long-term use of information systems. For the patient, as they are appraising stimuli against their ability to cope with a disease, therefore technology is considered one among many stimuli to the patient. Further, IQS involves the appraisal of the beliefs and attitudes associated with the information produced by a system, as a result of its use, as is planned to be received by the physician. Considering the persistent nature of IBD, and degree to which many IBD patients generally consult periodically with one physician (Bernstein et al., 2010), the ability for ESR tools to provide long-term support for both the patient and the physician, may be informed by integrating both perspectives into a multilevel model, with many patients nested within each physician. As this model is quite novel, it was necessary to consult several physicians and academics in its construction.

Choosing a Methodology

A sequential explanatory mixed methods approach will be employed (Creswell, 2009; Johnson, Onwuegbuzie & Turner, 2007), which is used when the objective of the research is to understand phenomena and to gain conceptual insight on a pre-existing model. As the model used in this research is new and unvalidated, there is certainly a need for gaining this insight. As well, the benefit of using a qualitative phase will inform the researchers more on the context being studied, and the "fit" of the technology to this context. This research will consist of two phases, a quantitative followed by a qualitative. In the quantitative phase, the impact of Gi BodyGuard as an ESR tool on the ability of the IBD patient to cope with their disease will be assessed. Participants will use and evaluate Gi BodyGuard in conjunction with their physicians. Quantitative data will be collected via an online questionnaire from participants before use at an appointment with their physician, and then again after their next regularly scheduled appointment, typically after one- to three months. In the qualitative phase, issues with including patient symptom data from Gi BodyGuard used as an ESR tool in clinical workflow will be explored. Qualitative interviews will be used to explore the determinant themes to facilitation or inhibition of long-term adherence to Gi BodyGuard as well as the ways in which data from Gi BodyGuard was used by the physician within their workflow. Results of the second phase will be used to update the model from the first phase, contributing to developing a general theory for ESR.

Recruitment of Participants

As the first phase of this study involves the use of a multilevel model integrating the perspectives of physician and patient, two populations must be defined, and two sets of sample frames sought. The physician population will be inclusive of physicians who care for IBD patients regularly, for example, gastroenterologists, family practitioners, as well as other relevant physicians. Nurse practitioners may also be included in this population. The sample frame will be sought from among membership listings, personal networks and events related to the Canadian Association of Gastroenterology, the Canadian Digestive Health Foundation, the Crohn's and Colitis Foundation of Canada, and the College of Family Practitioners of Canada. Physician participants will be self-selected; that is, they will volunteer. The sample will be further restricted to clinicians who are capable of communicating with their patients in English. To be eligible, physician participants will be screened based on the criteria described above. Physician participants must also be willing to recruit patients for the study in their clinic. Physicians that agree to participate will be asked to refer other potential physician participants. Patients that agree to participate will be compensated $100 for every patient they bring into the study.

The patient population will comprise individuals who have been diagnosed with a form of IBD, including but not strictly restricted to those diagnosed with CD and UC. As disease activity can change over the course of time, patients who are experiencing remission will also be permitted to participate. The sample frame is "defined English speaking patients who attend a participating physician clinic." As such, consenting and volunteering patients will be recruited through their physician offices. As this is a low-budget study, patients must also own a mobile device (e.g., an iPhone or Android device) capable of running Gi BodyGuard.

Retention of Participants

In order to retain physician participants, a separate strategy is formulated for physicians and patients, as their duties as participants in this study are quite different. As the physician is responsible for recruiting patients out of their own clinic, they must display the appropriate recruitment material as well as verbally attempt to recruit patients at their regular appointments. Either before or during patient follow up appointment, they will be expected to inspect the Gi BodyGuard data in conjunction with the patient, once for each patient they have successfully recruited in the study. They will then have to fill out their own questionnaire after the last patient has reported their results. As this involvement is substantial, and to encourage recruitment of many patients, the physician will be compensated with a fitting incentive of $100 per patient. They may also be selected for an in-depth interview, for the qualitative portion of the study, which will be compensated separately. As well, physicians will receive medical education credits as applicable. A rapport will be developed with the clinic staff as well as the physician, as they will need to prominently display the recruitment material, as well as schedule periodic telephone meetings and provide information when the physician is too busy. In the event that a clinic has not recruited participants in some while, a phone call will be arranged with the physician in order to assess any barriers to participation in the study experienced by patients. Physicians who are contacted that say that they wish to drop out of this study will be asked to participate in a brief telephone interview, concerning their barriers to participation in the study, their disuse of the

application, or any negative outcomes experienced. Physicians must explicitly drop out of the study before contact attempts will cease.

Patient involvement in the study consists of the following. They are to fill out a questionnaire at the start of the study, which will capture a majority of the data concerning their current disease activity, disease perception and coping strategies, perceptions of their physician, and perceptions to technology in general. They are then to download the Gi BodyGuard application on their phone, and use the application to log details concerning any disease activity or disease exacerbations, as well as undergo their regular routine. They are then to produce the data to the physician at their next regularly scheduled appointment. They are free to experiment with functions in the application, or to use the data themselves. After their appointment, they are to fill out another questionnaire, same as the first, but including questions on their perspective of the Gi BodyGuard application. They may also be selected for an in-depth interview, for the qualitative portion of the study, which will be compensated separately. In order to increase retention of the patient, their personal contact information, including email address and phone numbers, will be asked for at the beginning of the study. As well, the date of their next regularly scheduled appointment with their physician will be asked. Leading up to this date, the patient will be sent reminders to bring their symptom data into the appointment. After this date, patients will be sent reminders that they must fill out the final questionnaire. Patients who are contacted that say that they wish to drop out of this study will be asked to participate in a brief telephone interview, concerning their barriers to participation in the study, their disuse of the application, or any negative outcomes experienced. Patients must explicitly drop out of the study before contact attempts will cease. Participants that complete the final questionnaire will be compensated $100. If the rate of attrition is significant, a post hoc analysis will be conducted to see if any knowledge can be discovered from the existing data as to why participants dropped out of the study.

Creating Value for Developers

As future phases of Gi BodyGuard are planned (Mulvale, 2013), this formative evaluation is expected to inform the development team to current needs and future improvements. Although some validation of assumptions has occurred in the development of this technology concerning its use by IBD patients (Mulvale, 2013), this study is a formative evaluation within a clinical context. This study will serve to validate various assumptions the developers have in sharing symptom data from Gi BodyGuard with physicians.

Conveying Value to the Reviewers

As it is likely that reviewers from many areas related to eHealth will review this proposal, the purpose and value of the review must be communicated clearly. In particular, reviewers from the medical and healthcare areas view RCTs as the "gold standard" in research, therefore they may not be aware of the value of other types of research such as formative evaluations. In writing this research proposal, the researchers have communicated the value of formative evaluations in the following ways. First of all, the title of the research will explicitly contain the words "formative evaluation". To add to this, formative evaluations will be characterized in the body text of the proposal, along with describing the value of formative evaluations, their distinction from summative evaluations, and the reason why evaluations like this are performed at this time. As well, the contribution to the theory must be conveyed to the reviewers, as well as the developers in further developing this technology.

CONCLUSION

This article described the design of a formative evaluation of a smartphone application, applied for ESR within the context of IBD care. The goal of this application in the context of IBD care is to facilitate the recording and transmission of patient symptom data to healthcare providers, so that they may incorporate the data into their workflow. Theoretical challenges include integrating theories from outside of healthcare, and use of theories with insufficient empirical support. Methodological challenges include integration and utilization of exploratory methods into research projects. Practical challenges include producing value for developers, and communicating clearly the value of formative evaluations for reviewers that may not be familiar with this type of research.

REFERENCES

References available on request from Michael S. Dohan.