Résumé

Considérations méthodologiques sur l'évaluation des interventions en télésanté

Huong Q. Nguyen, DorAnne Cuenco, Seth Wolpin, Josh Benditt et Virginia Carrieri-Kohlman

Les progrès des technologies de l'information et de la communication façonnent les attitudes des consommateurs en ce qui a trait à leur engagement face à leur propre santé et leur interface avec le système de santé. L'utilisation d'outils de télésanté est très prometteuse, car elle favorise et facilite les changements en matière de comportements touchant la santé ainsi que la prévention et la gestion des maladies chroniques. Les auteurs se penchent sur des questions liées à la conception des études, à l'application des traitements et à la mesure des résultats dans le cadre des essais en télésanté, en fournissant des exemples issus de la littérature et de leurs propres études en cours. La sélection du groupe témoin et les considérations en matière de conception à partir des préférences des participants se fondent sur l'état des connaissances scientifiques et sur les pratiques courantes dans le domaine concerné. Une conception aléatoire permet de réduire les biais de sélection, et on la préfère dans les essais visant à évaluer l'efficacité potentielle et réelle des interventions en télésanté. Selon le choix des groupes témoins, les applications en télésanté doivent être passablement solides pour démontrer leur supériorité sur le traitement de référence. Les stratégies visant à assurer la fidélité au traitement et l'engagement soutenu des participants peuvent présenter des difficultés et ne réussissent pas toujours. Les résultats rapportés par les patients se retrouvent dans toutes les études sur la télésanté. Les autres résultats, comme les coûts associés aux nouvelles applications en télésanté, sont tout autant, sinon plus importants pour les décideurs. Cette discussion vise à éclairer les futurs essais et, de ce fait, à faire progresser les connaissances en télésanté.

Mots clés : télésanté, évaluation

Methodological Considerations in Evaluating eHealth Interventions

**Huong Q. Nguyen, DorAnne Cuenco, Seth Wolpin,
Josh Benditt, and Virginia Carrieri–Kohlman**

Advances in information and communication technologies shape consumers' attitudes towards engagement in their own health and their interface with the health-care system. The use of eHealth tools holds promise for supporting and enabling health behaviour change and the prevention and management of chronic diseases. The authors review issues related to study design, treatment implementation, and outcome measurement in eHealth trials, providing examples from the literature and from their own ongoing studies. Selection of a comparison group and design considerations related to participant preferences are based on the state of the science and current practice in the particular field. Randomized designs allow for control of selection bias and are favoured in both efficacy and effectiveness trials of eHealth interventions. Depending on the choice of comparison groups, eHealth applications must be fairly robust to demonstrate their efficacy above and beyond active controls. Strategies to ensure treatment fidelity and ongoing participant engagement can be challenging and are not always successful. Patient-reported outcomes are common to eHealth studies. Other outcomes, such as the costs associated with new eHealth applications, are equally if not more important for decision-makers. This discussion is intended to inform future trials and thereby serve to advance the science of eHealth.

Keywords: eHealth, Internet, telehealth, evaluation, methodology, dyspnea, chronic obstructive pulmonary disease

Introduction

Advances in information and communication technologies continue to shape consumers' attitudes towards engagement in their own health and their interface with the health-care system (Dickerson & Brennan, 2002). These changes, combined with a greater burden of chronic illness in the population, an aging population, and escalating health expenditures, have created significant challenges. Yet ample opportunities have emerged for nurses to explore innovative ways to positively impact the health of individuals and populations (Bodenheimer, 2005). In 2001, Eng surveyed the developing landscape of information and communication technologies in health care and coined the term eHealth, defining it as "the use of emerging interactive technologies (inclusive of all media types) to enable

health improvement and health care services" (Eng, 2001, p. 1). Other terms, such as consumer health informatics (Eysenbach, 2000) and inter-active behaviour change technology (Glasgow, Bull, Piette, & Steiner, 2004), emerged at the same time or soon thereafter. We will use eHealth to refer to information and communication technologies that directly engage health consumers and/or their families; discussions related to specific software applications to support health providers are beyond the scope of this article.

Notable among the eHealth tools designed to support health pro-motion and collaborative disease management are health-risk assessments linked to online and offline health promotion activities; applications that allow consumers to track and share health parameters; telehealth tech-nologies that support virtual home visits and ongoing monitoring; and decision-support tools ranging from automated expert systems to evidence-based practice guidelines and online peer communities that help health consumers and their families weigh risks and benefits asso-ciated with various treatments. The use of eHealth tools in general holds tremendous promise for the support and enabling of behaviour change and the prevention and management of chronic diseases. However, a recent survey of the eHealth landscape found that, although there has been progress in the field, high-quality evidence to support the efficacy and effectiveness of these applications is still limited (Ahern, Kreslake, & Phalen, 2006).

While the methodological challenges associated with clinical studies of eHealth applications are not substantively different from those asso-ciated with evaluations of nursing interventions that do not employ tech-nology, they can be amplified in unanticipated ways by the intersection between people and their use of technology. The purpose of this article is to review issues related to study design, treatment implementation, and outcome measurement in eHealth clinical trials and to illustrate these issues by highlighting examples from the literature and three of our ongoing studies. By bringing attention to these evaluation challenges and reporting on our "lessons learned," we hope to inform future trials and thereby help to advance the science of eHealth.

Study Designs

As imperfect as it may be, we will use the clinical trials framework of Phase I-IV, which guides pharmaceutical research as a heuristic to illus-trate the sequential steps associated with building the evidence base for eHealth applications (US Food and Drug Administration). In the context of drug trials, new pharmaceutical compounds often undergo early testing in healthy volunteers in Phase I to determine optimal intensity,

timing, mode of action or mechanism, tolerability, and safety. Similarly, short-term one- or two-group pre/post-test designs may be appropriate to determine feasibility, uncover usability and logistical issues, and establish early evidence of efficacy for trials of eHealth applications. Phase II is typically an explanatory clinical trial to determine whether a drug has the expected impact on a primary outcome. At this stage, randomized controlled trials (RCTs) with optimal testing conditions — for example, highly selected and motivated participants — are used to determine the efficacy of an eHealth application.

Once efficacy is established, the drug is administered to a larger group of participants to confirm its effectiveness, monitor side effects, and compare it to commonly used treatments (Phase III). eHealth interventions that reach Phase III are subjected to practical or pragmatic trials. These studies continue to employ randomization and are controlled, usually for community or standard alternative interventions with fewer eligibility restrictions so that the question of usefulness and generalizability can be properly addressed. Individual preferences are considered within the study design and tailored treatment algorithms (Glasgow, Davidson, Dobkin, Ockene, & Spring, 2006; Tunis, Stryer, & Clancy, 2003). Phase IV post-marketing or surveillance studies, which often employ observational cohort designs, are conducted after drugs are approved and are in broad use. Since eHealth tools are not regulated or required to undergo any formal approval process, their rapid diffusion into practice can present notable challenges to this formulaic sequential testing strategy. The framework described above informs our ensuing discussion regarding the selection of comparison groups and use of random assignment, two equally difficult and interdependent study design decisions.

Selection of Comparison Group(s)

The selection of a comparison group for studies testing a new eHealth intervention warrants serious consideration and often rests on the state of the science and practice in an area. In health care, there is frequently a standard of care against which new treatments can be compared. It should be noted that usual care is not necessarily comparable to standard care, especially since it has been found that patients receive, on average, only 50% of recommended standard care (McGlynn et al., 2003). The former question of most interest in eHealth research is whether a technology-enhanced intervention is comparable to or better than an existing standard intervention. Studies to address this question, often referred to as non-inferiority, equivalence, or comparative trials, require greater attention to study design, conduct, analysis, and interpretation. A new eHealth intervention may be expected to match the efficacy of the

standard treatment but have advantages in convenience or cost or simply be an alternative (Jones, Jarvis, Lewis, & Ebbutt, 1996). Non-inferiority designs are also appropriate for situations in which the prospect of not offering a viable treatment to all participants raises ethical concerns (Ellenberg & Temple, 2000; Temple & Ellenberg, 2000).

Various comparison conditions have been used in eHealth studies. These include usual care (Barnason, Zimmerman, Nieveen, & Hertzog, 2006), face-to-face counselling (Carlbring et al., 2005; Izquierdo et al., 2003), Web-based discussion groups (Andersson et al., 2005), and computerized assessment (Glasgow, Nutting, et al., 2004). Unfortunately, the majority of studies do not provide an explicit rationale for their selection of comparison groups. In some areas, such as smoking cessation, for which a clear standard of care exists, the studies of new eHealth tools are always compared to active behavioural interventions (Strecher, Shiffman, & West, 2005; Vidrine, Arduino, Lazev, & Gritz, 2006). In others, such as back pain, where the condition is often self-limiting and there is no clear effective therapy, usual-care controls are used (Lorig et al., 2002; Polly, 2005). Because standard of care differs across diseases and care settings and can change over time, effect sizes across eHealth studies are not always comparable and will become increasingly difficult to interpret (Murray, Burns, See, Lai, & Nazareth, 2005). The expected outcome differences between two active treatments are considerably smaller than those between active treatment and usual care.

Too often, latter-phase studies of eHealth interventions fail to perform rigorous head-to-head comparisons with standard treatments. The result is that stakeholders are left without crucial information about the relative cost-effectiveness, benefit, or risk of the new, competing treatments. An example of where an equivalence trial would have been useful is a recent study that tested the effects of an Internet-based chronic disease self-management program against usual care (Lorig, Ritter, Laurent, & Plant, 2006). The Internet program was adapted from a face-to-face program that had been widely disseminated and was considered the standard of care for patients with various chronic conditions. The new program had only modest effects on patient-reported outcomes and no impact on health-service use. The authors went on to conduct exploratory comparisons with a *historical* cohort that previously completed the face-to-face intervention and concluded that the results were similar for the two programs.

Before discussing our decision-making on a comparison group for the design of our current eHealth studies of dyspnea self-management in patients with chronic obstructive pulmonary disease (COPD), a brief overview of our work leading up to this stage is in order. Our cumulative work in understanding the symptom and coping strategies began

with early descriptive studies (Carrieri & Janson–Bjerklie, 1986; Carrieri, Janson–Bjerklie, & Jacobs, 1984), followed by two randomized trials of variations in the face–to–face dyspnea self–management program (Carrieri-Kohlman et al., 2005; Carrieri-Kohlman, Gormley, Douglas, Paul, & Stulbarg, 1996). This was followed by a Phase I study of the Internet-based dyspnea self-management program (eDSMP) using a one-group pre/post-test design (Nguyen, Carrieri-Kohlman, Rankin, Slaughter, & Stulbarg, 2005). Our DSMPs provide patient education and training on strategies for dyspnea management with a strong emphasis on independent exercise and ongoing reinforcement by a nurse.

For our first efficacy study of the eDSMP (Dyspnea Self-Management Study I), we proposed to compare the effects of the eDSMP to our "gold standard" face–to–face program (fDSMP) on the primary outcome of dyspnea with activities of daily living. We chose the fDSMP as a comparison treatment for two reasons. First, practice guidelines at the time recommended education for patients with COPD (National Heart, Lung and Blood Institute and World Health Organization, 2003). We also anticipated that the standard of care for patients with COPD would improve in the next few years and ultimately wished to build the evidence base for informed patient choice with the two programs. Second, this design provided a more stringent test of treatment efficacy and was perceived as an improvement over other eHealth studies where usual care was the comparison condition (Gustafson et al., 2001; Lorig et al., 2002; Southard, Southard, & Nuckolls, 2003).

Although our rationale for designing Dyspnea Self-Management Study I was conceptually sound subject to funding constraints, a three-arm design, which includes an attention control group, would (1) help guard against the potential of falsely concluding that the eDSMP is as efficacious as the fDSMP and that both were better than nothing, and (2) allow testing against the non-specific effects of attention, which we had not done before (Friedman, Furberg, & DeMets, 1998). In our expanded follow-up study, Dyspnea Self-Management Study II, we added a parallel control group that will receive general health education and comparable contact time. Some may view it as unethical to not provide these participants with information specific to their condition. Our rationale is that, despite national practice recommendations, structured education and skills training to manage COPD are still not the standard of care in primary care settings (Mularski et al., 2006). Moreover, participants assigned to this group will continue to have unrestricted access to online information resources available to any Internet user. Participants in the control group receive an intervention (initial home visit, six group-based health-education sessions, and ongoing telephone contact) that mimics the fDSMP, not the eDSMP.

Randomize or Accommodate Participant Preferences (or Both)?

The randomized experimental design, a cornerstone of evidence-based practice, provides the strongest evidence on the *efficacy* of eHealth interventions. It has been argued that randomized trials may be vulnerable to "preference effects" such that treatment preference and concordance could influence treatment adherence and consequently health outcomes (Sidani, 2004). With the trend towards greater patient participation in health-care decisions and the increased importance of patient-centred care (Davis, Schoenbaum, & Audet, 2005; Institute of Medicine, 2001), it has been argued that patient preferences play a key factor in determining the success of medical and behavioural treatments (Bradley, 1993; Brewin & Bradley, 1989). Efforts to foster patient involvement have been encouraged as a means to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcomes (Fisher et al., 2005; TenHave, Coyne, Salzer, & Katz, 2003). Wait-list control and group RCT designs have also been proposed by some researchers to better accommodate individual preferences while retaining randomization. However, these designs have their own limitations — for example, unacceptable treatment delays with wait-list designs and a large number of sites required for group RCTs (Gross & Fogg, 2001).

While there are hybrid designs that combine randomization and preference schemes to preserve causal inferences (Janevic et al., 2003; Noel et al., 1998), a simple preference trial that allows participants to select their treatment arm *is* a prospective observational study with its inherent susceptibility to confounding and selection biases. Partial RCT designs with a parallel preference cohort have serious problems in settings where participants have strong preferences for one treatment over another. In one study that employed such designs, an overwhelming 90% of participants selected the preference arm, leaving 10% in the randomized study (Miranda, 2004). Analyses of preference trials are inherently difficult and reports from these trials have been mixed and difficult to interpret (Bedi et al., 2000; Ward et al., 2000; Weinstein et al., 2006). Although results from these studies show that outcomes were not different for those who received their preferred treatment and those who were randomized, none of the studies was sufficiently powered to test the choice by treatment-group interaction (Bedi et al., 2000; Janevic et al.; Noel et al., 1998; Ward et al.).

Another possible alternative is a matched controlled design where random assignment is not used. Patients who choose to use an eHealth tool are compared to those who choose not to. Statistical techniques such as propensity scoring (PS) are employed to adjust for treatment selection bias in these designs. The use of PS is the observational study analogue of

randomization in RCTs (D'Agostino, 1998). A PS is a measure of the probability that a patient will receive an intervention. It is usually derived from a regression analysis of that patient's observed baseline characteristics. The PS essentially summarizes these characteristics in a single composite variable. Patients can then be stratified or matched based on this score and their outcomes compared within strata of comparable likelihood of receiving the eHealth intervention (Earle et al., 2001; Penrod et al., 2006). Although PS adjustments can balance observed baseline covariates between intervention and comparison groups, they do nothing to balance unobserved characteristics and confounders.

The US Department of Veterans Affairs has been in the forefront of testing technology-enhanced home-care services to better serve geographically distributed veterans. Propensity scoring was used to match patients who received a telecare management intervention to comparison participants (Barnett et al., 2006; Chumbler, Vogel, et al., 2005). Although health-service use was reported to be lower in the telecare intervention after 12 months compared to the matched controls, we cannot be sure if this reduction was due to the intervention itself or to some other factor that was not accounted for in the propensity-score adjustments. In another study, daily monitoring was compared to weekly monitoring for veterans with diabetes. Patients self-selected their monitoring schedule. Health-resource use was significantly lower for participants who received daily monitoring. However, baseline differences between the two groups on a number of parameters suggested marked residual confounding, which can only be addressed with better adjustments or randomization (Chumbler, Neugaard, Ryan, Qin, & Joo, 2005).

In keeping with our research question in Study I, we chose to randomize participants to one of the two treatment arms. We sought to describe treatment preferences with our sample and plan to test its association with adherence and outcomes when the study is completed. At baseline, after participants were already informed of their group assignment, they are asked about their preferences during the initial face-to-face consultation. Of the 39 participants enrolled thus far, approximately 38% said they preferred assignment to eDSMP while similar proportions preferred the fDSMP (31%) or did not have a program preference (31%). Although the data set is incomplete and perhaps underpowered, we found significant associations between treatment preferences, age, and Internet use. Participants who preferred the eDSMP ($n = 15$) were younger, 62.1 ± 8.8, compared to those who preferred the fDSMP ($n = 12$), 70.5 ± 6.5, or had no preference, 72.5 ± 6.6 ($n = 12$), $p = .002$. Similarly, participants who preferred the eDSMP spent more hours per week on the Internet, 31.9 ± 24.2, compared to 8.9 ± 6.6 and 8.8 ± 4.8 for those who preferred the fDSMP or had no preference,

respectively, *p* < .001. Of note, our analyses of these first 39 randomized participants showed comparable baseline demographic characteristics and health status for the eDSMP (*n* = 21) and fDSMP (*n* = 18) arms.

If this study had been designed as a simple preference trial, we could not be certain about the degree to which age and Internet use, and also other, unmeasured, factors, would confound the findings. Younger and more experienced computer users might have differential uptake of the intervention and consequently have changes in their dyspnea that could not be fully attributed to the intervention itself. These observations lend support to our decision to use random assignment in this efficacy trial as our best measure against selection bias. No other technique has the same power to control for selection as a threat to internal validity (Cook & Campbell, 1979). Once the efficacy of eDSMP and fDSMP are established, we can proceed with a Phase III study to determine the effectiveness of intervening with patients using both modalities. Fixed adaptive designs (TenHave et al., 2003), which accommodate patient preferences for treatments with known efficacy, have been used in a number of late-phase pharmaco-behavioural and health-services trials (Activity Counseling Trial Writing Group, 2001; Berkman et al., 2003; Diabetes Prevention Program, 1999; Unutzer et al., 2002); none of these studies included an eHealth component, mainly because many technology-supported interventions are still in early-phase efficacy testing.

Treatment Implementation: Strengthening Effects and Minimizing Crossovers

Maintenance of treatment fidelity is important for all research trials. The overall goal of enhanced treatment fidelity is to increase scientific confidence that the changes in outcomes are attributable to the treatment by reducing random and unintended variability (Bellg et al., 2004). Fidelity is even more critical in the context of non-inferiority trials, since treatment crossovers will make the groups more similar than they actually are. Treatment fidelity and participant non-adherence need to be addressed in the context of the overall study design. Gross and Fogg (2001) argue persuasively that random assignment is inconsistent with the current thrust of health-care consumerism and the expectation of personal choice in matters of health and health care. They go on to suggest that the democratic balance of knowledge and power with the baby boom generation in particular will likely lead consumers to reject any efforts that restrict their options or control. These dire forecasts are more pertinent for treatments that people can access outside of a research protocol than for those interventions that are available only within a study. Unfortunately, since many technology-supported cognitive-

behavioural interventions for health promotion and disease management are readily accessible to patients, treatment fidelity can be easily compromised.

The challenges we face in communicating with our eDSMP and fDSMP participants are a good illustration of this issue. Because our research question is focused on the efficacy of the Internet as a "delivery channel" for dyspnea education and ongoing support for behaviour change, we have strategies in place to encourage participants in the eDSMP and the fDSMP to use e-mail and telephone, respectively, to communicate with the study nurse. The nurse also provides weekly and biweekly reinforcements through these two communication channels. We have provisions in the protocol to communicate with eDSMP participants via telephone should an urgent or complex matter arise. Although the nurse redirects participants by responding via e-mail or telephone according to their group assignment, a number of participants have inevitably "crossed over" and used both e-mail and telephone during the course of the study. It should be noted that although all of our participants use a computer on a regular basis, some have very clear preferences for communicating via telephone and have deliberately not answered study e-mails. The opposite has occurred with the fDSMP participants but with less frequency. These observations may be unique to our older sample; however, they raise the question of whether random assignment to a communication modality is sensible when people have preferences and will use the modality they are comfortable with.

Early studies of "stand alone" Internet-based psycho-educational interventions had disappointing results. Approximately 50% of those who initially signed up for the programs dropped out or never logged back in (Andersson, Stromgren, Strom, & Lyttkens, 2002; Clarke et al., 2002; Eysenbach, 2005). Programs that integrated live counsellors who corresponded with participants and received regular feedback were more successful (Tate, Jackvony, & Wing, 2003). In our study, all participants receive an initial face-to-face consultation with the study nurse, as a way to strengthen the eDSMP and increase participants' commitment to the study.

It is important to note that eHealth tools targeted at relatively healthy participants for general health promotion can be successfully implemented outside of existing care relationships. However, this may not be the case for tools designed to support chronic disease management. Studies of outsourced chronic disease management often report communication challenges with patients' health-care providers (Southard et al., 2003). We have faced similar situations in our attempts to facilitate more prompt therapy when participants report sustained worsening of their respiratory symptoms. When Shea and colleagues (2006) conducted a

Phase III study of diabetes case management for older Medicare beneficiaries using a sophisticated telemedicine system, they achieved only modest improvements in diabetes outcomes compared to usual care. The less than robust changes were somewhat expected, since the technology-enhanced case management was performed externally, not by a nurse or care manager from the practice where patients received their existing care. It is critical that testing of new eHealth interventions for patients with chronic illnesses be conducted in the context of existing care relationships.

Other implementation issues of potential import for eHealth studies include use of a "run-in" period and strategies to maintain participant engagement. In another ongoing study that tests the effects of a PDA-mediated exercise-persistence intervention for older adults with COPD (MOBILE — Mobilizing Support for Long Term Exercise), we included a 2-week run-in for prospective participants. Since this is a Phase I efficacy trial, we wanted to ensure that participants felt comfortable with the technology before they were randomized to one of the two treatment arms. Technology-supported health interventions are no different from other treatments in that novelty effects quickly wear off. Similar to findings reported elsewhere, our pilot study of the eDSMP showed a precipitous drop-off in Web-site use after the first month (Nguyen et al., 2005; Tate et al., 2003). Other researchers have used strategies such as monetary and material incentives to encourage ongoing participation (Bowling et al., 2006). Monetary incentives, even in the form of raffles or lotteries, need to be carefully crafted so as not to be perceived as coercive. Participants in our studies are issued PDA devices with data service, which allows them to access the Web for news, weather, and so forth; however, anecdotal reports suggest that, for the older cohort, these features have little appeal and are infrequently used. As with any intervention study, more creative strategies for encouraging participant engagement in eHealth interventions need to be explored.

Measuring Outcomes

Selection of outcomes in studies of technology-supported interventions is based on the most anticipated effects of the intervention, taking into account those outcomes of greatest relevance to both patients and decision-makers. Many eHealth studies use Web-based questionnaires, either singly or with other methods such as telephone and mail, to measure patient-reported outcomes (PRO) — for example, symptoms, health-related quality of life (HRQL), and satisfaction. The validity of Web-based questionnaires is no longer questioned, though it was when eHealth first emerged (Dillman, 2000; Ritter, Lorig, Laurent, &

Matthews, 2004). The science of Web-survey methodology continues to be refined (Couper, 2005). One of the exciting research areas with PRO measurement is the development of item banks and computerized-adaptive tests (CAT) for measuring symptoms and HRQOL domains (Kosinski, Bjorner, Ware, Sullivan, & Straus, 2006; Reeve, 2006). Since questions are tailored for each respondent with CAT, this approach to outcome assessment promises to increase efficiency and measurement precision, two factors that can potentially reduce the number of patients needed for clinical trials of eHealth applications without sacrificing statistical power.

Paper diaries have traditionally been the primary means of capturing outcomes or processes related to patient experiences in their daily lives. However, recent experimental findings confirm researchers' early suspicions that participants do not always adhere to diary protocols and that diary "hoarding" is common practice (Lauritsen et al., 2004; Stone, Shiffman, Schwartz, Broderick, & Hufford, 2003). In one study, electronic time-stamp records indicated that "compliance" with paper diaries was only 11% (Stone et al.). Given this knowledge, would it be appropriate to continue to use paper diaries for an active control arm in a study testing the effects of an eHealth intervention that includes an electronic diary device? If data validity is of primary import and all participants are issued devices, would the introduction of such technologies contaminate the control arm?

We were more interested in the impact of real-time data transmission on early intervention for COPD exacerbations in the eDSMP, and therefore continued to use paper diaries for the fDSMP in both of our dyspnea self-management studies. Our early experience with using a PDA/phone device to capture real-time symptom and exercise data from participants in Study I showed a mean response time of 22 hours from when queries were sent to when our servers received the data (Nguyen, Wolpin, Chiang, Cuenco, & Carrieri, 2006). We believe the delays in data submission were primarily due to the cumbersome vendor-supported hardware and software over which we had limited control. We have since developed our Web-based platform, replaced the device, and reconfigured our queries to be more parsimonious.

In contrast to other latter-phase eHealth studies that strictly rely on patient self-report and therefore can have national and even international reach (Lorig et al., 2006; Strecher et al., 2005), our efficacy studies include in-person assessments. We confirm participants' self-report of COPD with pulmonary function tests to reduce misclassification and for ongoing safety monitoring. Other studies that include daily pulmonary function monitoring as a core intervention component have given patients spirometers for self-testing at home (Finkelstein, Cabrera, &

Hripcsak, 2000). Thus, it is possible to increase the reach to target populations by modifying the data-collection approach, but this will need to be balanced with additional costs and/or testing reliability. For performance-based outcomes that require in-person testing, collaboration with existing research or practice networks may be a solution.

Economic outcomes are of considerable interest to payers and policy-makers and ideally are included in Phase III effectiveness trials of eHealth interventions (Tunis et al., 2003). A few studies have examined the cost impact of virtual home visits compared to traditional home care, with mixed success, mainly due to small sample sizes and other methodological weaknesses (Finkelstein, Speedie, & Potthoff, 2006; Hopp et al., 2006; Noel, Vogel, Erdos, Cornwall, & Levin, 2004); well-conducted parallel cost-effectiveness studies require significant resources. One such economic analysis will be forthcoming from a study testing the effects of a telecare diabetes-management intervention (Shea et al., 2006).

Conclusions

Although preliminary work has been done to promote the development of standards for evaluating eHealth applications targeted to health consumers (Eng, Gustafson, Henderson, Jimison, & Patrick, 1999), this article adds to the literature by delineating and examining in depth several key methodological challenges related to testing these technologies. We have reviewed issues related to study design, treatment implementation, and outcome assessment and have provided examples from the literature and our current work to illustrate some of the challenges. We have offered a few possible though imperfect solutions.

The design of any clinical trial begins and ends with the research question and trials of eHealth interventions are no exception. Thorough testing and evaluation of eHealth interventions require the use of many research questions, approaches, and designs. Selection of a clinically meaningful treatment alternative or comparison group and design considerations related to participant preferences will depend on the state of the science and current practice in the particular field. Nevertheless, randomized controlled designs are favoured when establishing the efficacy and effectiveness of eHealth applications. Depending on the choice of comparison groups, eHealth applications will need to be fairly robust to demonstrate their efficacy above and beyond active controls. Ensuring treatment fidelity and ongoing participant engagement can be particularly challenging and will require more creative solutions. And while more efficient methods of obtaining patient-reported outcomes are emerging, attention to other important outcomes, such as cost, will be critical in future evaluations of eHealth tools.

The advances made in information and communication technologies have undoubtedly ushered in renewed hope for and promise that such tools will facilitate positive changes in individual and population health. However, these are only promises and as such will require critical and systematic evaluations that carefully consider key methodological challenges.

References

Activity Counseling Trial Writing Group. (2001). Effects of physical activity counseling in primary care: The Activity Counseling Trial: A randomized controlled trial. *Journal of the American Medical Association, 286*(6), 677–687.

Ahern, D. K., Kreslake, J. M., & Phalen, J. M. (2006). What is eHealth (6): Perspectives on the evolution of eHealth research. *Journal of Medical Internet Research, 8*(1), e4.

Andersson, G., Bergstrom, J., Hollandare, F., Carlbring, P., Kaldo, V., & Ekselius, L. (2005). Internet-based self-help for depression: Randomised controlled trial. *British Journal of Psychiatry, 187,* 456–461.

Andersson, G., Stromgren, T., Strom, L., & Lyttkens, L. (2002). Randomized controlled trial of Internet-based cognitive behavior therapy for distress associated with tinnitus. *Psychosomatic Medicine, 64*(5), 810–816.

Barnason, S., Zimmerman, L., Nieveen, J., & Hertzog, M. (2006). Impact of a telehealth intervention to augment home health care on functional and recovery outcomes of elderly patients undergoing coronary artery bypass grafting. *Heart and Lung, 35*(4), 225–233.

Barnett, T. E., Chumbler, N. R., Vogel, W. B., Beyth, R. J., Qin, H., & Kobb, R. (2006). The effectiveness of a care coordination home telehealth program for veterans with diabetes mellitus: A 2-year follow-up. *American Journal of Management Care, 12*(8), 467–474.

Bedi, N., Chilvers, C., Churchill, R., Dewey, M., Duggan, C., Fielding, K., et al. (2000). Assessing effectiveness of treatment of depression in primary care: Partially randomised preference trial. *British Journal of Psychiatry, 177,* 312–318.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443–451.

Berkman, L. F., Blumenthal, J., Burg, M., Carney, R. M., Catellier, D., Cowan, M. J., et al. (2003). Effects of treating depression and low perceived social support on clinical events after myocardial infarction: The Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) randomized trial. *Journal of the American Medical Association, 289*(23), 3106–3116.

Bodenheimer, T. (2005). High and rising health care costs. Part 1: Seeking an explanation. *Annals of Internal Medicine, 142*(10), 847–854.

Bowling, M. J., Rimer, B. K., Lyons, E. J., Golin, C. E., Frydman, G., & Ribisl, K. M. (2006). Methodologic challenges of e-health research. *Evaluation and Program Planning, 29,* 390–396.

Bradley, C. (1993). Designing medical and educational intervention studies: A review of some alternatives to conventional randomized controlled trials. *Diabetes Care, 16*(2), 509–518.

Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomised clinical trials. *British Medical Journal, 299*(6694), 313–315.

Carlbring, P., Nilsson-Ihrfelt, E., Waara, J., Kollenstam, C., Buhrman, M., Kaldo, V., et al. (2005). Treatment of panic disorder: Live therapy vs. self-help via the Internet. *Behavior Research and Therapy, 43*(10), 1321–1333.

Carrieri, V. J., & Janson-Bjerklie, S. (1986). Strategies patients use to manage the sensation of dyspnea. *Western Journal of Nursing Research, 8,* 284–305.

Carrieri, V., Janson-Bjerklie, S., & Jacobs, S. (1984). The sensation of dyspnea: A review. *Heart and Lung, 13*(436–437).

Carrieri-Kohlman, V., Gormley, J. M., Douglas, M. K., Paul, S. M., & Stulbarg, M. S. (1996). Exercise training decreases dyspnea and the distress and anxiety associated with it: Monitoring alone may be as effective as coaching. *Chest, 110*(6), 1526–1535.

Carrieri-Kohlman, V., Nguyen, H. Q., Donesky-Cuenco, D., Demir-Deviren, S., Neuhaus, J., & Stulbarg, M. S. (2005). Impact of brief or extended exercise training on the benefit of a dyspnea self-management program in COPD. *Journal of Cardiopulmonary Rehabilitation, 25*(5), 275–284.

Chumbler, N. R., Neugaard, B., Ryan, P., Qin, H., & Joo, Y. (2005). An observational study of veterans with diabetes receiving weekly or daily home telehealth monitoring. *Journal of Telemedicine and Telecare, 11*(3), 150–156.

Chumbler, N. R., Vogel, W. B., Garel, M., Qin, H., Kobb, R., & Ryan, P. (2005). Health services utilization of a care coordination/home-telehealth program for veterans with diabetes: A matched-cohort study. *Journal of Ambulatory Care Management, 28*(3), 230–240.

Clarke, G., Reid, E., Eubanks, D., O'Connor, E., DeBar, L. L., Kelleher, C., et al. (2002). Overcoming Depression on the Internet (ODIN): A randomized controlled trial of an Internet depression skills intervention program. *Journal of Medical Internet Research, 4*(3), e14.

Cook, T., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field setting.* Boston: Houghton-Mifflin.

Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review, 23*(4), 486–501.

D'Agostino, R. B., Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistical Medicine, 17*(19), 2265–2281.

Davis, K., Schoenbaum, S. C., & Audet, A. M. (2005). A 2020 vision of patient-centered primary care. *Journal of General Internal Medicine, 20*(10), 953–957.

Diabetes Prevention Program. (1999). Design and methods for a clinical trial in the prevention of type 2 diabetes. *Diabetes Care, 22*(4), 623–634.

Dickerson, S. S., & Brennan, P. F. (2002). The Internet as a catalyst for shifting power in provider-patient relationships. *Nursing Outlook, 50*(5), 195–203.

Dillman, D. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: John Wiley.

Earle, C. C., Tsai, J. S., Gelber, R. D., Weinstein, M. C., Neumann, P. J., & Weeks, J. C. (2001). Effectiveness of chemotherapy for advanced lung cancer in the elderly: Instrumental variable and propensity analysis. *Journal of Clinical Oncology, 19*(4), 1064–1070.

Ellenberg, S. S., & Temple, R. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. *Annals of Internal Medicine, 133*(6), 464–470.

Eng, T. R. (2001). *The eHealth landscape*. Princeton, NJ: Robert Wood Johnson Foundation.

Eng, T. R., Gustafson, D. H., Henderson, J., Jimison, H., & Patrick, K. (1999). Introduction to evaluation of interactive health communication applications: Science Panel on Interactive Communication and Health. *American Journal of Preventive Medicine, 16*(1), 10–15.

Eysenbach, G. (2000). Consumer health informatics. *British Medical Journal, 320*(7251), 1713–1716.

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1), e11.

Finkelstein, J., Cabrera, M. R., & Hripcsak, G. (2000). Internet-based home asthma telemonitoring: Can patients handle the technology? *Chest, 117*(1), 148–155.

Finkelstein, S. M., Speedie, S. M., & Potthoff, S. (2006). Home telehealth improves clinical outcomes at lower cost for home healthcare. *Telemedicine Journal and E-Health, 12*(2), 128–136.

Fisher, E. B., Brownson, C. A., O'Toole, M. L., Shetty, G., Anwuri, V. V., & Glasgow, R. E. (2005). Ecological approaches to self-management: The case of diabetes. *American Journal of Public Health, 95*(9), 1523–1535.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials*. New York: Springer.

Glasgow, R. E., Bull, S. S., Piette, J. D., & Steiner, J. F. (2004). Interactive behavior change technology: A partial solution to the competing demands of primary care. *American Journal of Preventive Medicine, 27*(2 Suppl), 80–87.

Glasgow, R. E., Davidson, K. W., Dobkin, P. L., Ockene, J., & Spring, B. (2006). Practical behavioral trials to advance evidence-based behavioral medicine. *Annals of Behavioral Medicine, 31,* 5–13.

Glasgow, R. E., Nutting, P. A., King, D. K., Nelson, C. C., Cutter, G., Gaglio, B., et al. (2004). A practical randomized trial to improve diabetes care. *Journal of General Internal Medicine, 19*(12), 1167–1174.

Gross, D., & Fogg, L. (2001). Clinical trials in the 21st century: The case for participant-centered research. *Research in Nursing and Health, 24*(6), 530–539.

Gustafson, D. H., Hawkins, R., Pingree, S., McTavish, F., Arora, N. K., Mendenhall, J., et al. (2001). Effect of computer support on younger women with breast cancer. *Journal of General Internal Medicine, 16*(7), 435–445.

Hopp, F., Woodbridge, P., Subramanian, U., Copeland, L., Smith, D., & Lowery, J. (2006). Outcomes associated with a home care telehealth intervention. *Telemedicine Journal and E-Health, 12*(3), 297–307.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington: National Academy Press.

Izquierdo, R. E., Knudson, P. E., Meyer, S., Kearns, J., Ploutz-Snyder, R., & Weinstock, R. S. (2003). A comparison of diabetes education administered through telemedicine versus in person. *Diabetes Care, 26*(4), 1002–1007.

Janevic, M. R., Janz, N. K., Dodge, J. A., Lin, X., Pan, W., Sinco, B. R., et al. (2003). The role of choice in health education intervention trials: A review and case study. *Social Science and Medicine, 56*(7), 1581–1594.

Jones, B., Jarvis, P., Lewis, J. A., & Ebbutt, A. F. (1996). Trials to assess equivalence: The importance of rigorous methods. *British Medical Journal, 313*(7048), 36–39.

Kosinski, M., Bjorner, J. B., Ware, J. E., Jr., Sullivan, E., & Straus, W. L. (2006). An evaluation of a patient–reported outcomes found computerized adaptive testing was efficient in assessing osteoarthritis impact. *Journal of Clinical Epidemiology, 59*(7), 715–723.

Lauritsen, K., Degl' Innocenti, A., Hendel, L., Praest, J., Lytje, M. F., Clemmensen-Rotne, K., et al. (2004). Symptom recording in a randomised clinical trial: Paper diaries vs. electronic or telephone data capture. *Controlled Clinical Trials, 25*(6), 585–597.

Lorig, K. R., Laurent, D. D., Deyo, R. A., Marnell, M. E., Minor, M. A., & Ritter, P. L. (2002). Can a back pain e-mail discussion group improve health status and lower health care costs? A randomized study. *Archives of Internal Medicine, 162*(7), 792–796.

Lorig, K. R., Ritter, P. L., Laurent, D. D., & Plant, K. (2006). Internet-based chronic disease self-management: A randomized trial. *Medical Care, 44*(11), 964–971.

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine, 348*(26), 2635–2645.

Miranda, J. (2004). An exploration of participants' treatment preferences in a partial RCT. *CJNR, 36*(3), 100–114.

Mularski, R. A., Asch, S. M., Shrank, W. H., Kerr, E. A., Setodji, C. M., Adams, J. L., et al. (2006). The quality of obstructive lung disease care for adults in the United States as measured by adherence to recommended processes. *Chest, 130*(6), 1844–1850.

Murray, E., Burns, J., See, T. S., Lai, R., & Nazareth, I. (2005). Interactive health communication applications for people with chronic disease. *Cochrane Database System Review, 4,* CD004274.

National Heart, Lung and Blood Institute and World Health Organization. (2003). *NIH National Heart, Lung and Blood Institute Global Initiative for Chronic Obstructive Lung Disease: Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease Workshop report.* Publication # 2701. Washington: National Institutes of Health.

Nguyen, H. Q., Carrieri-Kohlman, V., Rankin, S. H., Slaughter, R., & Stulbarg, M. S. (2005). Is Internet-based support for dyspnea self-management in patients with COPD possible? Results of a pilot study. *Heart and Lung, 34*(1), 51–62.

Nguyen, H. Q., Wolpin, S., Chiang, K. C., Cuenco, D., & Carrieri, K. V. (2006). Exercise and symptom monitoring with a mobile device. *AMIA Annual Symposium Proceedings,* 1047.

Noel, H. C., Vogel, D. C., Erdos, J. J., Cornwall, D., & Levin, F. (2004). Home telehealth reduces healthcare costs. *Telemedicine Journal and E-Health, 10*(2), 170–183.

Noel, P. H., Larme, A. C., Meyer, J., Marsh, G., Correa, A., & Pugh, J. A. (1998). Patient choice in diabetes education curriculum: Nutritional versus standard content for type 2 diabetes. *Diabetes Care, 21*(6), 896–901.

Penrod, J. D., Deb, P., Luhrs, C., Dellenbaugh, C., Zhu, C. W., Hochman, T., et al. (2006). Cost and utilization outcomes of patients receiving hospital-based palliative care consultation. *Journal of Palliative Medicine, 9*(4), 855–860.

Polly, D. W. (2005). An Internet-delivered cognitive-behavioral intervention with telephone support improved some coping skills in patients with chronic low back pain. *Journal of Bone and Joint Surgery, 87*(5), 1169.

Reeve, B. B. (2006). Special issues for building computerized-adaptive tests for measuring patient-reported outcomes: The National Institute of Health's investment in new technology. *Medical Care, 44*(11 Suppl 3), S198–204.

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research, 6*(3), e29.

Shea, S., Weinstock, R. S., Starren, J., Teresi, J., Palmas, W., Field, L., et al. (2006). A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus. *Journal of the American Medical Informatics Association, 13*(1), 40–51.

Sidani, S. (2004). Rethinking the research-practice gap: Relevance of the RCT to practice. *CJNR, 36*(3), 7–18.

Southard, B. H., Southard, D. R., & Nuckolls, J. (2003). Clinical trial of an Internet-based case management system for secondary prevention of heart disease. *Journal of Cardiopulmonary Rehabilitation, 23*(5), 341–348.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials, 24*(2), 182–199.

Strecher, V. J., Shiffman, S., & West, R. (2005). Randomized controlled trial of a Web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction, 100*(5), 682–688.

Tate, D. F., Jackvony, E. H., & Wing, R. R. (2003). Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: A randomized trial. *Journal of the American Medical Association, 289*(14), 1833–1836.

Temple, R., & Ellenberg, S. S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Annals of Internal Medicine, 133*(6), 455–463.

TenHave, T. R., Coyne, J., Salzer, M., & Katz, I. (2003). Research to improve the quality of care for depression: Alternatives to the simple randomized clinical trial. *General Hospital Psychiatry, 25*(2), 115–123.

Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association, 290*(12), 1624–1632.

Unutzer, J., Katon, W., Callahan, C. M., Williams, J. W., Jr., Hunkeler, E., Harpole, L., et al. (2002). Collaborative care management of late-life depression in the primary care setting: A randomized controlled trial. *Journal of the American Medical Association, 288*(22), 2836–2845.

Vidrine, D. J., Arduino, R. C., Lazev, A. B., & Gritz, E. R. (2006). A randomized trial of a proactive cellular telephone intervention for smokers living with HIV/AIDS. *Aids, 20*(2), 253–260.

Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., et al. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *British Medical Journal, 321*(7273), 1383–1388.

Weinstein, J. N., Lurie, J. D., Tosteson, T. D., Skinner, J. S., Hanscom, B., Tosteson, A. N., et al. (2006). Surgical vs. nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT) observational cohort. *Journal of the American Medical Association, 296*(20), 2451–2459.

Authors' Note

Comments or queries may be directed to Huong Q. Nguyen, University of Washington, HSB T602A, Box 357266, Seattle, Washington 98199 USA. Telephone: 206-543-8651. Fax: 206-543-4771. E-mail: HQN@u.washington.edu.

Huong Q. Nguyen, RN, PhD, is Assistant Professor, School of Nursing, University of Washington, Seattle, United States. DorAnne Cuenco, RN, PhD, is Assistant Adjunct Professor, School of Nursing, University of California, San Francisco, United States. Seth Wolpin, RN, PhD, is Research Assistant Professor, School of Nursing, University of Washington, Seattle. Josh Benditt, MD, is Professor, School of Medicine, University of Washington, Seattle. Virginia Carrieri-Kohlman, RN, DNSc, is Professor, School of Nursing, University of California, San Francisco.