Title
Continuous Evaluation of Evolving Behavioral Intervention Technologies

Authors
Mohr, David C
Cheung, Ken
Schueller, Stephen M
et al.

Continuous Evaluation of Evolving Behavioral Intervention Technologies

David C. Mohr, PhD, **Ken Cheung, PhD**, **Stephen M. Schueller, PhD**, **C. Hendricks Brown, PhD**, and **Naihua Duan, PhD**

Department of Preventive Medicine and the Center for Behavioral Intervention Technologies (Mohr, Schueller), Northwestern University, Chicago, Illinois; the Department of Biostatistics (Cheung) and Department of Psychiatry (Duan), Columbia University, New York City, New York; and the Department of Epidemiology and Public Health (Brown), University of Miami, Miami, Florida

Abstract

Behavioral intervention technologies (BITs) are web-based and mobile interventions intended to support patients and consumers in changing behaviors related to health, mental health, and well-being. BITs are provided to patients and consumers in clinical care settings and commercial marketplaces, frequently with little or no evaluation. Current evaluation methods, including RCTs and implementation studies, can require years to validate an intervention. This timeline is fundamentally incompatible with the BIT environment, where technology advancement and changes in consumer expectations occur quickly, necessitating rapidly evolving interventions. However, BITs can routinely and iteratively collect data in a planned and strategic manner and generate evidence through systematic prospective analyses, thereby creating a system that can "learn."

A methodologic framework, Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT), is proposed that can support the evaluation of multiple BITs or evolving versions, eliminating those that demonstrate poorer outcomes, while allowing new BITs to be entered at any time. CEEBIT could be used to ensure the effectiveness of BITs provided through deployment platforms in clinical care organizations or BIT marketplaces. The features of CEEBIT are described, including criteria for the determination of inferiority, determination of BIT inclusion, methods of assigning consumers to BITs, definition of outcomes, and evaluation of the usefulness of the system. CEEBIT offers the potential to collapse initial evaluation and postmarketing surveillance, providing ongoing assurance of safety and efficacy to patients and consumers, payers, and policymakers.

Introduction

Behavioral intervention technologies (BITs) are web-based and mobile interventions intended to support patients and consumers in changing behaviors related to health, mental health, and well-being.[1] The term BIT is used, rather than eHealth, as eHealth is a broad term encompassing areas such as health informatics.[2] Although some BITs may be designed for single or short-term use, such as a decision-making tool, many are designed for longer-termuse aimed at supporting sustained behavior change.[3] BITs are deployed individually or

---

as part of larger deployment platforms. They may be made available directly to consumers through commercial marketplaces, patient portals, or may be "prescribed" by care providers.

A growing interest globally is in developing robust deployment platforms, which would ideally be able to interact with electronic medical records.[4] For example, in the U.S., the Veterans Administration is developing and implementing a mobile application deployment platform to be integrated with the "My HealtheVet" portal.[5] In England, the Institute of Psychiatry together with the South London and Maudsley National Health Service Trust is developing a BIT deployment platform available through a patient portal. The aim of the current paper is to propose a methodologic framework, Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT), that can support the rapid evaluation of BITs in near–real time through deployment sites located in care-providing organizations or commercial marketplaces with the aim of protecting consumers from ineffective or inferior BITs.

Behavioral intervention technologies are being developed and used at a rapidly increasing rate. For example, a recent report noted more than 97,000 mobile health applications, with the top 10 generating up to 4 million free and 300,000 paid downloads per day.[6] Evaluation of efficacy or effectiveness of the vast majority of these BITs is lacking.[7] Evaluation is critical to protect the public from harmful or ineffective BITs, to support the development of models that can integrate BITs into the healthcare and reimbursement systems, to support decision making by consumers and providers, and to protect the credibility of the emerging BIT industry.

Although the FDA has drafted guidelines for mHealth medical devices, this regulation will likely not include consumer applications.[8] Other services, such as Beacon (beacon.anu.edu.au/) provide information on the published evidence for BIT efficacy, and Happitique.com is planning to introduce an app certification program that focuses on operability, privacy, and security but does not evaluate efficacy. The evidence base remains weak, and virtually no discussion addresses *how* to evaluate BITs effectively and efficiently.

Current evaluation methodologies for behavioral interventions typically involve development, pilot testing, evaluation in at least two RCTs, and implementation studies.[9,10] Estimates are that it can take up to 17 years from initial research to full implementation for medical and behavioral interventions.[11–13] This timeline is fundamentally incompatible with the BIT environment, where technology advancement and changes in consumer expectations happen quickly, necessitating rapidly evolving interventions. Obsolescence can occur because of either technical advances in which a newer device supersedes an older one (e.g., the replacement of personal digital assistants by smartphones) or through sudden shifts in the technology environment (the rapid decline of the Nokia Symbianmobile operating system in the U.S. and the rise of Android). Technological development, improved computing power, and greater capacity for data transmission continually increase our capacity to develop new intervention tools.[14] Consumers have come to anticipate such change and can be intolerant of BITs that do not meet their expanding expectations. Thus, the shelflife of BITs may be shorter than the time it takes to evaluate a BIT using current methodologies and standards.[15]

Given the rapid pace of development, a recent NIH Workshop on "mHealth Evidence" called for alternatives to RCTs.[15] The CEEBIT methodology offers such an alternative. Using data generated by BITs, CEEBIT can evaluate multiple interventions or evolving versions of interventions in an "open-panel horserace," eliminating "horses" that demonstrate poorer outcomes, while allowing new horses to enter at any time. The horserace does not have to end and declare a winner, as long as innovations continue to emerge and

interventions continue to evolve; the objective is to have better and better horses running as time goes on, thereby providing patients with better and better care. This paper provides an overview of the proposed methodology, describes criteria for the elimination of a BIT, discusses potential randomized and nonrandomized methods of assigning consumers to BITs, addresses change in consumer population over time, reviews outcome evaluation, and considers evaluation of deployment systems using CEEBIT.

Methods

Overview

The CEEBIT strategy is proposed as a method of evaluation that would integrate data from multiple BITs provided by one deployment system (e.g., a clinical care organization or commercial marketplace) that would (1) acquire outcome and use data from BITs; (2) perform analytics to identify inferior BITs; (3) allocate consumers to BITs; and (4) remove inferior BITs from the system. An example is provided to illustrate the evaluation of the sequential introduction of new BITs targeting a single clinical outcome within a single deployment system (Figure 1). BIT A is deployed into the field at Time 1. Subsequent BITs are deployed at Times 2, 3, and 4. All versions are maintained until a version meets an a priori criterion for inferiority and is eliminated.

In the illustration, BIT A is determined to be inferior to BIT D at Time 5 and thus eliminated. Similarly, BIT C is determined to be inferior to BIT D and eliminated at Time 6, and BIT B is determined to be inferior to BIT D at Time 7, leaving BIT D. Versions are retained in the deployment system and are available to consumers until they have demonstrated inferiority. The inferiority criterion is used for BITs, rather than superiority, as requiring that a BIT demonstrate superiority to all other BITs would expose greater numbers of consumers to inferior BITs. By continually removing inferior BITs, the overall system should produce observable improvements in outcomes in the consumer population outcomes. Below, the considerations and specifications of the CEEBIT methodology are reviewed.

Determination of Inclusion

A CEEBIT administrator, in consultation with clinicians and BITs experts, would make decisions regarding the entry of new BITs into the deployment system. These decisions would be made based on a variety of criteria, which could include information from prior evaluations conducted during development, pilot testing, trials, or if such information is lacking, based on expert opinion. Evaluation could be initiated with a comparison to a control arm, which might consist of a generic intervention, to allow the assessment for efficacy/effectiveness. However, it is likely that most care-provider organizations would be more interested in using a comparative-effectiveness paradigm, evaluating BITs with similar treatment goals that have met some minimal threshold for entry into the system.

Determination of Inferiority

Inferiority should be determined by meeting a statistical criterion that an application produces less improvement on the target symptom, behavior, or outcome criterion relative to another application or applications in the system. The conventional RCT design tests a single new intervention against a standard using a small Type I error (i.e., alpha=0.05 and high statistical power [i.e., 0.80]), often resulting in trials of several hundred patients. If the statistical analyses demonstrate that the newer intervention has significantly improved outcomes compared to the standard, this single trial is used as evidence in support of altering future treatment.

In the context of comparing two care-delivery procedures in a single treatment setting, Cheung and Duan[16] have proposed a twostage strategy that compares the procedures in an investigation stage and "rolls out" the successful procedure to the remaining population in the consumption stage. Applying the conventional 5%–80% rule to determine the sample size in the investigation stage of two or more applications would require a large sample size that would limit feasibility and utility in the fast-changing landscape of BITs. Larger numbers of patients required for the investigation stage would reduce the number of patients who would "consume" the results of the trial and increase the number of patients exposed to the inferior treatment.

Based on cost-effectiveness considerations for both the investigation and consumption stages, the optimal design rule corresponds to a much more liberal Type I error rate than 5%. Specifically, under equipoise, the optimal design that maximizes the expected net gain corresponds to a 50% Type I error rate.[16] Additionally, the use of a 50% Type I error rate formulates the comparative study as a selection problem that assumes *symmetry of preference* among the tested arms.[17] The 5% Type I error rate used in conventional hypothesis testing implies a preference for the standard treatment arm; that is, when the statistical test result is insignificant, the standard arm is "retained." However, if no standard BIT exists to be tested against, the selection paradigm (which corresponds to a 50% Type I error) is a relevant and appropriate perspective that has been shown to be applicable in the context of clinical studies.[18] Importantly, as a consequence of this more liberal Type I error rate, the required sample size is considerably reduced.

Requiring inferiority to all BITs would under most circumstances be an unnecessarily high bar, and would do little to protect consumers. Figure 1 displays elimination based on inferiority to at least one other BIT. This would be the least-conservative approach to protecting consumers from BITs that have demonstrated any evidence of inferiority. If the remaining BITs were assumed to be equivalent until inferiority to one application is demonstrated, using this method of removal would not reduce the overall efficacy of the deployment system, and would provide some measureable improvement. Less-conservative decision rules might also be employed when multiple BITs are available, such as eliminating a BIT if it is found to be inferior to two or more BITs, or to a pooled average of two or more BITs. Further, a BIT may demonstrate inferiority for the population as a whole, but be effective within subgroups. In this case, adjustments to the methodology could be made to include subgroups in both deployment algorithms and evaluation methodologies.

Outcomes

Within CEEBIT, two primary types of outcomes can be used to determine inferiority: *clinical outcomes* and *use* (or adherence). Clinical outcome data represent the treatment target, such as weight, smoking cessation, or depression. These outcomes would generally use patient self-reports collected through the application, but they could also include clinician reports and clinical data entered via an electronic medical record. The time frame for outcome evaluation within a user would be set by the CEEBIT administrator, depending on the treatment target, and potentially also observed use patterns.

No accepted standard definition of BIT application use exists[19]; use has been measured in various ways, including the frequency that a consumer has logged in or used a program, components of a program that have been used, and length of time on an application. Outcome and use data can be partially independent. Although use is a necessary but not sufficient prerequisite for a BIT to have an effect, the relationship between use and clinical outcome is not clear and would likely vary by application and clinical target.[20] Use statistics

may be better markers for internal evaluation, whereas clinical outcomes may be better suited for program evaluation.

Managing Evolution of the Technologies

Behavioral intervention technologies are continuously evolving through frequent programming updates. When a substantial change in functionality occurs, administrators may choose to reintroduce the BIT de novo into the deployment system. For small changes (i.e., bug fixes or cosmetic changes), it might be reasonable to treat the new variant as a continuation of the original variant, but weighting data based on time to gradually focus on the new variant. Weighting can be handled through the Bayesian framework for combining data for the new variant with data from the original variant. For large changes that are systemic and structural, an administrative decision to evaluate the new variant on its own without considering the original variant may be required.

Assignment of Consumers

Consumers often want to be aware of treatment options and to have choice when multiple options are available.[21] Forcing consumers to accept one BIT when they prefer another could result in decreased motivation to use the application, thereby biasing the results[22] and decreasing the utility and acceptability of a deployment system. Thus, methods of including consumer choice and characteristics in the assignment of treatments are necessary and described below, both with and without randomization. Each of the assignment methods described below has advantages and disadvantages. This is one area of the CEEBIT methodology that will require further investigation and optimization.

Randomized methods—Preference-based randomization allows consumer preferences to be identified prior to treatment assignment. Those who express a preference for one BIT can receive their preference, whereas those who identify no preference or more than one option can be randomized to acceptable BITs.[22,23] Consumer choice can be incorporated into the evaluation as a fully observed pre-randomization factor. Contrasts among randomized consumers permit comparison of the BIT efficacy among similar consumers. Contrasts between randomized and nonrandomized subjects may indicate the additional influence of choice. If acceptable to consumers, preference-based randomization may be an optimal balance between preserving consumer choice and optimizing CEEBIT analytics.

However, for consumers to understand their choices, they would have to review descriptions, examples, or previews of each option, after which preferences could be enumerated. As the number of choices expands, this may become overly burdensome. A "randomize, reject, reassign" method, used in some pharmaceutical trials,[24] would randomize consumers to a currently available BIT. If dissatisfied, the consumer would have the option of rejecting the BIT and would then be randomly reassigned to another BIT. Rejection of an option would count as a failure for the rejected option. The "randomize, reject, reassign" method may reduce the burden of reviewing all BITs; however, it may also result in growing frustration and decreased motivation as the number of rejected BITs grows; the likelihood of rejection may also increase as consumers hope to find the perfect BIT.

Additional protection may be offered by using response-adaptive randomization (RAR), in which the randomization probabilities to different BITs may be unbalanced adaptively, based on interim data, to favor assignment to the BIT having empirically superior outcomes. This is ethically appealing because more patients are assigned to the more-successful BITs. In a Bayesian inference context, the randomization probability of a given BIT can be computed as the posterior probability that this given BIT is superior to the others,[25] although

non-Bayesian formulation is also available (e.g., randomized play-the-winner[26] and sequential elimination).[17,27] This approach can also incorporate patient covariates, including preference, in the randomization,[28] to address patient heterogeneity.

Nonrandomized assignment methods—Although randomization is usually considered the gold standard for the design of evaluation studies, nonrandomized assignment methods are warranted in situations in which randomization is not feasible because of ethical or practical constraints. A common concern of nonrandomized assignment methods is potential selection bias arising from associations between assignment and observed or unobserved prognostic factors, which can confound outcomes and prognostic factors.[29]

Various statistical methods have been developed to mitigate selection bias in nonrandomized studies. For example, the propensity score method can mitigate overt bias when all confounding factors are observed.[29–31] This method balances patients across treatment arms with respect to observed covariates to measure the effect of the intervention on the outcomes. Another promising nonrandomized assignment method is need-based assignment (NBA), also called risk-based assignment or regression-discontinuity design.[32–34] With NBA, patients are prescreened at baseline with a measure of need and assigned to an appropriate intervention. Patient outcomes are analyzed using the regression discontinuity model that includes one segment of regression function for patients with baseline need below the threshold and another segment for patients with baseline need above the threshold.

Contamination—It is possible, and indeed likely, that outcomes may be contaminated by the use of services or BITs outside the deployment system, which could enhance or reduce clinical and use outcomes. As some BITs encourage outreach, it is possible that BITs may have differential effects on health-seeking behaviors and access of services. The impact on health-seeking behaviors, however, could be considered part of the effect of the BIT, and therefore would not necessarily represent true contamination. The validity of the system could be compromised if consumers with a greater propensity to access outside resources select specific interventions.

Such biases could be introduced from various sources. For example, consumers with a greater propensity to seek care may do more research to select BITs that are better suited to them. Alternatively, physicians may differentially recommend applications to patients based on factors related to the patient's ability or desire to access services. In the absence of methods to comprehensively evaluate the consumer's capacity and propensity to seek services, randomized methods of assignment are preferable.

Stationarity: Managing Changing Populations and Environments

Consumer populations and cultures, policies, and practices related to BITs are likely to change over time, violating the statistical assumption of stationarity (that samples are constant over the evaluation period) and making earlier results less relevant. These changes can be incorporated into CEEBIT models by decreasing the weighting of older data. For example, as the devices through which intervention applications are delivered achieve greater penetration and are used by more-diverse populations, characteristics of these newer groups may incrementally affect the outcomes of interest.

Currently, mobile health applications are more likely to be used by African Americans and Latinos than Caucasians, but less likely to be used by the elderly.[35] Thus, as elderly consumers' access increases, overall outcomes may change based on that group's comfort and propensity to use health applications. Alternatively, the populations served by the

deployment system may change suddenly. For example, a care setting may acquire new clinics serving different types of potential consumers. If such changes in the population are detected or anticipated, their effects on clinical and use outcomes can be monitored and potentially mitigated by including baseline variables believed to identify such shifts as covariates in the CEEBIT models.

Evaluation of a Deployment System

It will be important to be able to track the usefulness of the CEEBIT methodology in improving deployment systems over time. Initially, the concept will have to be evaluated, both to test its efficacy (program evaluation) and to make refinements to improve its performance (quality improvement and internal evaluation). If effective, the CEEBIT methodology, which aims to improve the quality of BIT applications deployed within a care system or marketplace, should demonstrate improvements in outcomes among the population of consumers over time as a result of the elimination of inferior BITs and the introduction and retention of superior BITs. In real-world implementation, the benefits of the system would also have to include evaluation of cost effectiveness.

If stationarity can be assumed, a longitudinal model (either parametric such as a linear time trend model, or nonparametric, depending on the plausibility of the parametric model and the richness of the data) can be fit to the use and outcome data to ascertain whether an increasing trend exists that indicates benefit attributable to the CEEBIT framework. This evaluation is focused on the performance for the overall framework, instead of the performance of individual BITs. Evaluation of the overall system is analogous to the evaluation of FDA regulations regarding the discovery of beneficial new drugs and the elimination of harmful would-be drugs. If stationarity cannot be assumed, say, if the composition of the client pool evolves over time, potential confounding factors, such as baseline clinical severity and/or markers of technologic literacy, can be assessed and adjusted for in the longitudinal model to ensure the validity for the findings.

Cost effectiveness is also a critical outcome in evaluating the value of the CEEBIT framework. Methods of costing will need to be developed that would include the cost of system development; maintenance (e.g., updates and bug fixes to manage changes in the technologic environment); prevention of obsolescence; hosting and administrative costs; and costs to the consumer or payor of the BIT. Incremental costs would likely decline as more consumers are added to the system.

Conclusion

Continuous evaluation of evolving BITs harnesses the capacity to collect outcome and use data in real time, and subject it to continuous evaluation to evaluate and compare BITs. Evaluation frameworks, such as CEEBIT, could prove valuable to all stakeholders. The public would have reliable assurance of the efficacy and safety of BITs. Developers would gain analytic methodologies to evaluate the usefulness of changes they make to interventions and indicate opportunities for improvement. Policymakers would have information that would support decision making.

Because CEEBIT operates in real time, it can accommodate the rapidly changing technologic environment and would be responsive to changing expectations of consumers. New interventions, or new "evolved" BITs versions, can be inserted at any time into a deployment system along with existing interventions for real-time comparison, eliminating those that prove to be inferior. This eliminates the need for multiple individual RCTs, which may or may not be representative of real-world conditions at the trial's conclusion.

One might argue that the proposed CEEBIT methodology exposes consumers to BITs that have not been evaluated and that may be ineffectual or harmful. The current authors would argue that consumers are already exposed to such risks, and that CEEBIT provides an efficient method of weeding out less-effective BITs. Perhaps this is not ideal, but under the current conditions, it is unlikely that a methodology can perfectly protect consumers through data collected prior to the release of a BIT.

Continuous evaluation of evolving behavioral intervention technologies eliminates the distinction between initial evaluation and postmarketing surveillance. Currently, devices (as well as pharmaceuticals) are monitored after their approval to refine, confirm, or deny safety in the population at large. Such surveillance requires the integration of data from various sources including spontaneous reporting databases, patient registries, and health records. CEEBIT allows continuous monitoring of the efficacy and safety of BITs within populations, and can do so relative to other available interventions.

Continuous evaluation of evolving behavioral intervention technologies represents a first proposal for such a methodology, which could be deployed broadly in care-providing and commercial marketplaces. As a first proposal, many areas require increased specification, such as the need to develop methods of comparing the same outcomes assessed with different measures, and the fact that many applications have multiple and partially overlapping aims (e.g., a mixture of weight loss, healthy diet, physical activity). Additionally, the potential to integrate BIT databases to provide even more powerful evaluation tools is not discussed. Nevertheless, CEEBIT does capitalize on data generated by BITs to continuously evaluate efficacy in a manner consistent with the current sociotechnologic environment. CEEBIT has the potential to provide needed information to consumers and other stakeholders that would enhance safety and decision making.

Acknowledgments

References

1. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry. 2013; 35(4):332–338. [PubMed: 23664503]

2. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res. 2005; 7(1):e1. [PubMed: 15829471]

3. Fogg BJ, Hreha J. Behavior wizard: a method for matching target behaviors with solutions. Persuasive Technol Proc. 2010; 6137:117–131.

4. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med. 2013; 10(2):e1001382. [PubMed: 23424286]

5. Houston, TK.; Frisbee, K.; Moody, R., et al. Medicine 2.0. Boston, MA: 2012. Implementation of Veterans Affairs (VA) Mobile Health (Panel).

6. research2guidance. Mobile health market report 2013–2017: the commercialization of mHealth applications (Vol. 3). Berlin: research2-guidance; 2013.

7. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med. 2011; 40(3):279–285. [PubMed: 21335258]

8. U.S. Food and Drug Administration. Draft guidance for industry and Food and Drug Administration staff—mobile medical applications. 2011. www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm263280.htm

9. Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol. 1998; 66(1):7–18. [PubMed: 9489259]

10. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. Clin Psychol Science Pract. 2001; 8(2):133–142.

11. Balas, EA.; Boren, SA. Managing clinical knowledge for healthcare improvements. In: van Bemmel, JH.; McCray, AT., editors. Yearbook of medical informatics 2000. Vol. 2000. Stuttgart: Schattauer; p. 65-70.

12. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health. 2003; 93(8):1261–1267. [PubMed: 12893608]

13. Brown CH, Kellam SG, Kaupert S, et al. Partnerships for the design, conduct, and analysis of effectiveness, and implementation research: experiences of the prevention science and methodology group. Adm Policy Ment Health. 2012; 39(4):301–316. [PubMed: 22160786]

14. Christensen, CM.; Johnson, CW.; Horn, MB. Disrupting class: how disruptive innovation will change the way the world learns. New York: McGraw Hill; 2008.

15. Kumar S, Nilsen W, Abernethy A, et al. Mobile health technology evaluation: the Mhealth evidence workshop. Am J Prev Med. 45(2):228–236. [PubMed: 23867031]

16. Duan N, Cheung YK. Design of implementation studies for quality improvement programs: an effectiveness/cost-effectiveness framework. Am J Pub Health. 2013 (In Press).

17. Levin B, Robbins H. Selecting the highest probability in binomial or multinomial trials. Proc Natl Acad Sci U S A. 1981; 78(8):4663–4666. [PubMed: 16593069]

18. Cheung YK, Gordon PH, Levin B. Selecting promising ALS therapies in clinical trials. Neurology. 2006; 67(10):1748–1751. [PubMed: 17130405]

19. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res. 2009; 11(2):e13. [PubMed: 19403466]

20. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res. 2011; 13(3):e52. [PubMed: 21821503]

21. Guadagnoli E, Ward P. Patient participation in decision-making. Soc Sci Med. 1998; 47(3):329–339. [PubMed: 9681902]

22. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. BMJ. 1989; 299(6694):313–315. [PubMed: 2504416]

23. Lavori PW, Rush AJ, Wisniewski SR, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. Biol Psychiatry. 2001; 50(10):792–801. [PubMed: 11720698]

24. Stroup TS, McEvoy JP, Swartz MS, et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. Schizophr Bull. 2003; 29(1):15–31. [PubMed: 12908658]

25. Thompson WR. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika. 1933; 25:285–294.

26. Wei LJ, Durham S. Randomized play—winner rule in medical trials. J Am Stat Assoc. 1978; 73(364):840–843.

27. Cheung YK. Simple sequential boundaries for treatment selection in multi-armed randomized clinical trials with a control. Biometrics. 2008; 64(3):940–949. [PubMed: 17970818]

28. Cheung YK, Inoue LY, Wathen JK, Thall PF. Continuous Bayesian adaptive randomization based on event times with covariates. Stat Med. 2006; 25(1):55–70. [PubMed: 16025549]

29. Rosenbaum, PR. Observational studies. New York NY: Springer; 2002.

30. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70(1):41–55.

31. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc. 1984; 79(387):516–524.

32. Finkelstein MO, Levin B, Robbins H. Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. Am J. Public Health. 1996; 86(5):696–705. [PubMed: 8629722]

33. Finkelstein MO, Levin B, Robbins H. Clinical and prophylactic trials with assured new treatment for those at greater risk: I. A design proposal. Am J Public Health. 1996; 86(5):691–695. [PubMed: 8629721]

34. West SG, Duan N, Pequegnat W, et al. Alternatives to the randomized controlled trial. Am J Public Health. 2008; 98(8):1359–1366. [PubMed: 18556609]

35. Smith A. Mobile access 2010: Pew Internet and the American Life Project. 2010
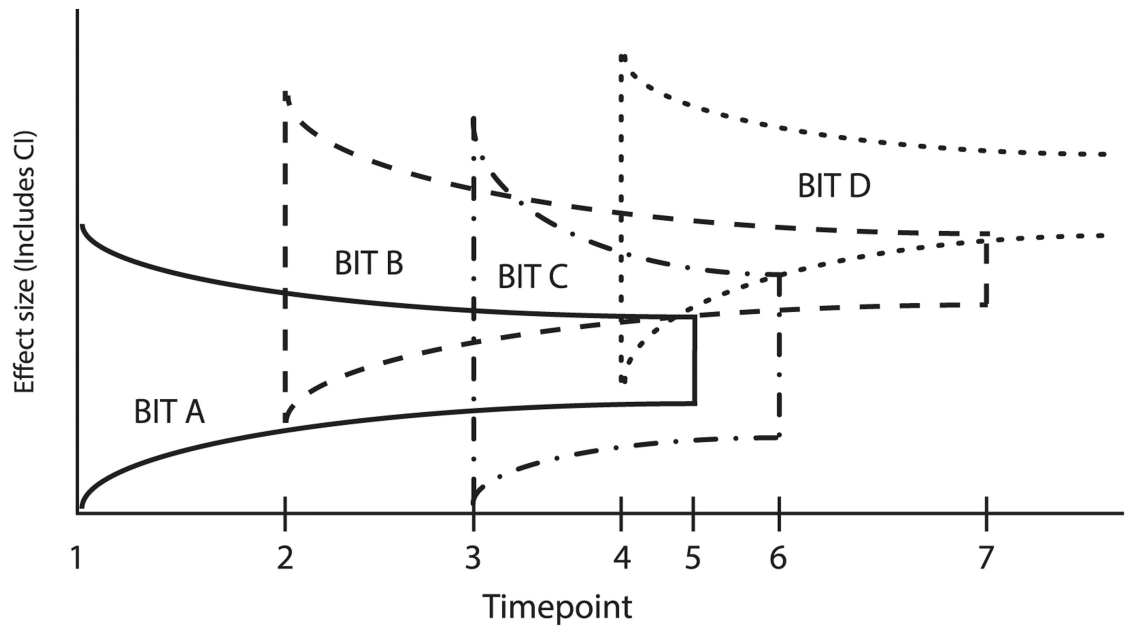
Figure 1.
Continuous evaluation of evolving BITs BIT, behavioral intervention technology