Evaluating Implementation Fidelity in Health Information Technology Interventions

Eric L. Eisenstein, D.B.A.1, David F. Lobach, M.D., Ph.D., M.S2, Paul Montgomery, D.Phil.3, Kensaku Kawamoto, Ph.D.2 and Kevin J. Anstrom, Ph.D.1
1Duke Clinical Research Institute, Duke University Medical Center, Durham, NC
2Department of Community & Family Medicine, Duke University Medical Center, Durham, NC
3Centre for Evidence Based Intervention, University of Oxford, Oxford, England

Health information technology evaluators need to distinguish between intervention efficacy as assessed in the ideal circumstances of clinical trials and intervention effectiveness as assessed in the real world circumstances of actual practice. Because current evaluation study designs do not routinely allow for this distinction, we have developed a framework for evaluation of implementation fidelity that considers health information technologies as complex interventions and makes use of common intervention components as defined in the Oxford Implementation Index. We also propose statistical methods for the evaluation of interventions at the system and component level using the Rubin Causal Model. We then describe how to apply this framework to evaluate an ongoing clinical trial of three health information technology interventions currently implemented in a 17,000 patient community-based health network caring for Medicaid beneficiaries in Durham County, North Carolina.

Introduction
Medical researchers have drawn distinctions between two types of tests for medical technology.1, 2 Tests of efficacy determine how a medical technology performs under ideal circumstances such as those observed in clinical trials; whereas, tests of effectiveness determine how the technology performs in usual circumstances such as those found in actual practice. Both types of tests answer important questions: tests of efficacy address the technical aspects of medical technologies and tests of effectiveness address the use of medical technologies in particular settings.

As clinical trials seek to be relevant to real world experiences, many of the distinctions between tests of efficacy and effectiveness have become blurred. The result is that when a new medical technology is found to be no better than its alternative, we may have no way of determining whether this negative result should be attributed to the technology itself or to the way it was implemented. This distinction is important as system developers need to know whether corrective actions should seek to improve the medical information technology itself or change the way it is used.

As an example, when patients fail to adhere to treatment regimens, they may suffer adverse outcomes that are related not to the efficacy of their treatments, but rather to the effectiveness of their use. While studies of non-adherence to evidence-based medical practices have largely focused on problems with physician prescription of appropriate treatments or patient compliance in following a treatment regimen, non-adherence has also been associated with the under-use of health information technologies.3, 4 Thus, the appropriate use of health information technologies appears to be another form of adherence that should be incorporated into evaluation study designs.

In the present study, we define a framework for evaluation of implementation fidelity that will allow evaluators to distinguish between and estimate the efficacy and effectiveness of health information technologies. We then test this framework by using it to design an implementation fidelity evaluation plan for an ongoing health information technology clinical trial.

Complex Interventions
While clinical trials traditionally have evaluated the effects of single interventions (such as a comparison of beta blocker A vs. beta blocker B), researchers are realizing that many interventions no longer fit this paradigm. Accordingly, the term complex interventions has been introduced to describe interventions involving several interconnected components.5 Many health information technology interventions fit into this category, whether they are evaluated as information systems per se (such as computerized decision support) or included as
components within a coordinated health care system (such as a heart failure clinic). Although various proposals have been advanced for the evaluation of complex interventions, there is general agreement that a careful understanding of the components of these interventions and their interrelationships is a prerequisite for effectiveness evaluation.\textsuperscript{5-7} It is also agreed that the overall complexity of an intervention may be a product of the complexity of the problem being addressed, the intervention being evaluated, inputs to and outputs from the health care domain, and the degree of user involvement.\textsuperscript{7} Obviously, if any of these components is not functioning properly, the effectiveness of the intervention may be diminished. We believe that previous research into implementation fidelity provides a conceptual framework for understanding key components of complex health information technology interventions.

**Implementation Fidelity: The Oxford Implementation Index**

It is generally recognized that some differences in the effectiveness of complex healthcare interventions may be attributable to variability in their administration.\textsuperscript{8} While there have been several attempts to develop and test methods for assessing implementation fidelity within specific domains (e.g., behavior change research), the Oxford Implementation Index is the first attempt to develop a general methodology that can be applied to many domains.\textsuperscript{9, 10} The Oxford Implementation Index identifies four implementation components that may affect treatment fidelity.\textsuperscript{9} These domains are: (1) treatment design, (2) treatment delivery by clinicians, (3) treatment uptake by patients, and (4) context factors. In this framework, treatment design includes factors such as the treatment setting, the intended treatment dose, other materials that are to be used in the intervention, and the training of clinicians and patients. Treatment delivery includes the actual dose administered, the quality and use of other materials, the qualifications of key staff members, and the use of quality management efforts to monitor adherence to protocols. Participant uptake includes all facets of the patient’s experience such as the core treatment components, the actual treatment dose received, and any non-protocol treatments received. Finally, context factors include all relevant aspects of the clinical study context such as inclusion/exclusion criteria for the patient population, patient and provider socioeconomic status, and other aspects of the health care environment such as managed care and type of insurance that may shape treatment delivery. The objective of the Oxford Implementation Index is to reveal “active components of a multi-component intervention, the contexts in which a treatment is most effective, or treatment components that cause harm.”\textsuperscript{9} We believe this index can be used to define key components of health information technology interventions and to suggest approaches for studies that evaluate implementation fidelity.

**Treatment Adherence: The Rubin Causal Model**

Statisticians have devised methods to determine the impact of non-adherence on clinical trial results.\textsuperscript{11-13} We will apply these methods to evaluate the efficacy and effectiveness of health information interventions, both for an entire system and within each of the implementation fidelity components. Sheiner and Rubin distinguish between use-effectiveness (what we term effectiveness) and method-effectiveness (what we term efficacy).\textsuperscript{13} These authors make the observation that while we would like to know what happens when someone receives a treatment (efficacy), clinical trial results only tells us what happens when a subject is assigned to a treatment group (effectiveness). The Rubin Causal Model (RCM) provides a framework for estimating the average causal effect of a treatment using information collected during a clinical trial.\textsuperscript{12} This information provides our best estimate of what happens when an individual receives a treatment. The RCM method is based upon the distinction between compliers (patients who will adhere to their assigned treatment) and never-takers (patients who will not adhere). Because patients are randomized to treatment assignment in a clinical trial, we assume that the proportion of compliers and never-takers will be identical in treatment groups, and that the observed effectiveness of a treatment will be the weighted average of that observed among compliers and never-takers.

<table>
<thead>
<tr>
<th>Study Arm</th>
<th>Adherence Distribution</th>
<th>Patient Event Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complier</td>
<td>75%</td>
<td>1%</td>
</tr>
<tr>
<td>Never-Taker</td>
<td>25%</td>
<td>20%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>5.7%</td>
</tr>
<tr>
<td><strong>Placebo</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complier</td>
<td>75%</td>
<td>10%</td>
</tr>
<tr>
<td>Never-Taker</td>
<td>25%</td>
<td>20%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>12.5%</td>
</tr>
</tbody>
</table>

In the example above, we evaluate a therapy designed to prevent adverse events among patients. There are two study arms (Treatment and Placebo) and the percent of patients who are compliers and never-takers is identical in both arms. We have observed a 6.8% absolute reduction (5.7% vs. 12.5%) in adverse events.
events among patients randomized to Treatment vs. Placebo. However, we know from Treatment results that 25% of these patients did not take the treatment (were never-takers), and that among these patients there was a 20% adverse event rate. After removing the never-takers from both study arms, we now have a 9% absolute reduction (10% vs. 1%) in adverse events among compliers randomized to the Treatment vs. Placebo arms. Thus, while the overall effectiveness of Treatment is sensitive to the proportion of never-takers (non-adherence) in our sample, the efficacy of Treatment (measured only in compliers) is not. This type of analysis allows us to separate the efficacy of our intervention from its effectiveness, which is in part a reflection of its use in a specific setting.

**Evaluation Framework for Implementation Fidelity**

By combining the Oxford Implementation Index and the Rubin Causal Model,9, 12 we are able to create a framework for evaluating implementation fidelity, with the Index defining the evaluation components and the Model providing the evaluation method (Figure 1). This framework will allow us to determine the efficacy and effectiveness of an entire health information technology intervention and of each implementation component (treatment design, treatment delivery, and treatment uptake). By performing these analyses, we will be able to determine the relative contributions of different components to the intervention’s overall effectiveness, and highlight specific components where improvements are needed.

![Figure 1: Framework for Evaluation of Implementation Fidelity](image)

**Problem Formulation**

Each HIT Value alert mechanism and its immediate sequels can be seen as a series of inter-connected events that take the form:

1. Priority health event occurs,
2. Alert is sent to clinician,
3. Clinician contacts patient,
4. Patient corrective health event is scheduled,
5. Patient completes corrective health event,
6. Priority health event does not recur.

For example, if a patient presents to the emergency room with an asthma diagnosis, an alert will be sent to the patient’s case worker. The case worker then is expected to contact the patient within 30 days and to arrange an asthma outpatient appointment for the patient to address presumed care management issues. The patient is expected to attend this appointment, and future emergency room visits with an asthma diagnosis for this patient are expected to decline.

However, a number of empirical questions remain. Do these events occur at the planned times? Does the variability in timing of these events impact the efficacy and effectiveness of the intervention? For each information intervention alert (email to case worker, letter to patient or parent, or report to primary care clinic), we have several points of
Table 2: Implementation Fidelity Analysis Framework Example

<table>
<thead>
<tr>
<th>Treatment Design</th>
<th>Delivery of Treatment</th>
<th>Patient Treatment Uptake</th>
<th>Contextual Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email Alert</td>
<td>Case Worker</td>
<td>Contact patient</td>
<td>Receive appropriate care</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Schedule appropriate</td>
<td>Not repeat alert behavior</td>
</tr>
<tr>
<td></td>
<td></td>
<td>care</td>
<td></td>
</tr>
<tr>
<td>Letter</td>
<td>Patient or Parent</td>
<td>Schedule appropriate</td>
<td>Receive appropriate care</td>
</tr>
<tr>
<td>Periodic Report</td>
<td>Patient’s Primary Care Clinic</td>
<td>Schedule appropriate care</td>
<td>Not repeat alert behavior</td>
</tr>
<tr>
<td>Standard of Care Group</td>
<td>None</td>
<td>Schedule appropriate care</td>
<td>Receive appropriate care</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Not repeat alert behavior</td>
</tr>
</tbody>
</table>

Table 2 illustrates the Oxford Implementation Index’s four components applied to the HIT Value study. In this example, four treatments are being studied. The information content of each active intervention is essentially the same; however, their designs differ in the form of transmission (email, letter, or report) and in their recipients (case workers, patients or their parents, and the patients’ primary care clinics). All analyses will be performed at the alert level and will include a subset of patients within the four treatment groups for whom a specific type of alert was triggered (e.g., low severity emergency department visit).

For each alert type, we will specify performance parameters that permit the estimation of adherence rates for clinicians, case workers and patients, and will allow us to determine how non-adherence impacts the study’s outcomes.

Clinician performance parameters relating to the delivery of treatment typically will include the scheduling of appropriate care (e.g., outpatient clinic visit containing a specific diagnosis code) within a pre-defined period of time from the alert. As we do not have access to clinic appointment schedules, we will not be able to track the scheduling of appropriate care. Case workers also will have performance parameters for patient contact.

Patient performance parameters relating to treatment uptake typically will include the receipt of appropriate care within a pre-defined period of time from the alert, as well as the absence of behavior which would prompt another alert for the same reason (the priority health event). This absence of subsequent alert prompting behavior is the primary study outcome for each alert.

From these performance parameters we will be able to identify and measure performance metrics. For the Delivery of Treatment metric, we will monitor case worker contact with patients in response to individual alerts. The information collected will include the type of patient contact (e.g., letter, phone call, home visit), and the time from alert transmission to patient contact. For alerts sent to case workers we will also capture the time interval between patient contact and receipt of appropriate care. For the Patient Treatment Uptake metric, we will collect the times from alert transmission to receipt of appropriate care (most often an outpatient visit with a specific diagnosis). Lastly, we will collect data on the recurrence of patient behavior that prompts the alerts.

All analyses will be performed using time-to-event as the primary outcome, and percent-of-events within 30-day time intervals as secondary outcomes. Comparisons will be made between each of the three active treatments and the standard of care, and also just between the active treatments. All analyses will be initially performed using the data as collected. These results will estimate the relative effectiveness of our three active information interventions. Next, the analyses will be repeated using the Rubin Causal Model. These results will estimate the relative efficacy of our three information interventions.

There are important limitations which will need to be addressed when using the implementation fidelity framework for empirical analyses. First, the RCM does not account for treatment concealment. Thus,
While contamination is clearly possible, the use of randomization at the household level in our study will serve to mitigate its occurrence. Second, the RCM does not identify the reasons for non-adherence. Although this information is not necessary to calculate treatment efficacy, it would be useful when designing corrective measures to know whether non-adherence was due to patient choice or to other factors which can be changed in subsequent implementations. For this reason, our study will interview samples of non-adherent clinicians and patients to determine their reasons for non-adherence.

**Discussion**

We have defined a framework for implementation fidelity evaluation which measures the efficacy and effectiveness of treatments. Through the use of this framework, researchers can determine whether system failures are due directly to information interventions or to problems with the methods of their implementation, and identify areas for improvement.

**Reference List**


