Health Services Research Evaluation Principles
Broadening a General Framework for Evaluating Health Information Technology

P. S. Sockolow1; P. R. Crawford2,3; H. P. Lehmann2,4,5
1College of Nursing and Health Professions, Drexel University, Philadelphia, PA, USA;
2Department of Health Sciences Informatics, School of Medicine, The Johns Hopkins University, Baltimore, MD, USA;
3Department of International Health, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA;
4Department of Health Policy and Management, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA;
5Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA

Keywords
Evaluation studies, technology evaluation, health information systems, medical informatics, review

Summary
Background: Our forthcoming national experiment in increased health information technology (HIT) adoption funded by the American Recovery and Reinvestment Act of 2009 will require a comprehensive approach to evaluating HIT. The quality of evaluation studies of HIT to date reveals a need for broader evaluation frameworks that limits the generalizability of findings and the depth of lessons learned.

Objective: Develop an informatics evaluation framework for health information technology (HIT) integrating components of health services research (HSR) evaluation and informatics evaluation to address identified shortcomings in available HIT evaluation frameworks.

Method: A systematic literature review updated and expanded the exhaustive review by Ammenwerth and deKeizer (AdK). From retained studies, criteria were elicited and organized into classes within a framework. The resulting Health Information Technology Research-based Evaluation Framework (HITREF) was used to guide clinician satisfaction survey construction, multi-dimensional analysis of data, and interpretation of findings in an evaluation of a vanguard community health care EHR.

Results: The updated review identified 128 electronic health record (EHR) evaluation studies and seven evaluation criteria not in AdK: EHR Selection/Development/Training; Patient Privacy Concerns; Unintended Consequences/ Benefits; Functionality; Patient Satisfaction with EHR; Barriers/Facilitators to Adoption; and Patient Satisfaction with Care. HITREF was used productively and was a complete evaluation framework which included all themes that emerged.

Conclusions: We can recommend to future EHR evaluators that they consider adding a complete, research-based HIT evaluation framework, such as HITREF, to their evaluation tools suite to monitor HIT challenges as the federal government strives to increase HIT adoption.

1. Introduction

Our forthcoming national experiment in increased health information technology (HIT) adoption funded by the American Recovery and Reinvestment Act of 2009 [1] will require a comprehensive approach to evaluating HIT. The quality of evaluation studies of HIT to date reveals a range of shortcomings that limit the generalizability of findings and the depth of lessons learned [2, 3]. For example, findings of the impact of a computerized physician order entry system on mortality rates that are contradictory [4, 5] cannot be compared due to differences in system implementation and evaluation-research design [6].

The informatics community has tended to adopt and adapt evaluation frameworks from the management information systems (MIS) literatures [7]. Currie has depicted HIT evaluation frameworks as describing “methodologies that capture the processes integral to applications, the users and the world in which the users function” [8, p. 909]. Some frameworks are models that describe the interrelations among variables (i.e., Technology Acceptance Model) [9] while other frameworks describe the relationship between a dimension of the framework and a result (i.e., measure progress toward adoption of HIT [10, 11]). Currie has suggested a classification of HIT evaluation frameworks consisting of four general groups. Frameworks are differentiated based on the principles used as a guide: behavioral; social; software development life cycle; or none of these (i.e., generic) [8]. Applying this classification to MIS frame-
works, generic frameworks include DeLone and McLean’s Information Systems Success (ISS) model which assesses the success of MIS using factors related to the impact of utilization and acceptance on individual performance [12]; and Goodhue and Thompson’s Task-Technology fit model which highlights the match between user tastes and needs, and the available IT functionality [13]. An example of behavioral-focused MIS evaluation framework is Technology Acceptance Model (TAM) which graphically represents computer usage behavior and user acceptance of IT [9].

In biomedical informatics, generic frameworks include those that evaluate progress towards developing and implementing health information infrastructure [10, 11]. Behavioral-focused frameworks have been proposed by Horsky et al. to evaluate the user interface of clinical ordering systems [14]. Examples of social-organizational focused frameworks include those of: Kaplan’s framework to assess HIT [15]; Winkelman and Leonard’s framework to assess patient use of electronic patient records [16]; Green and Moehr’s framework to assess performance of Canadian HIT [17]; Lau et al.’s framework to evaluate benefits of Canadian HIT [18]; Kukafka et al.’s model to identify factors that impede system usage behavior [19]; Holden and Karsh’s framework to develop and guide reporting systems [20] and the PRISM framework that assesses performance of routine health information systems in developing countries by incorporating environmental and behavioral determinants of performance [21]. The system development life cycle framework is exemplified by Stead et al.’s phased approach to evaluation that reflects the sequence of information-system development and implementation [22].

To date HIT evaluation studies have tended to focus on technology and to exclude contextual considerations (e.g., professional, organizational, social) that influence whether the system will be used or how it will be used [25]. Recent review of 12 HIT evaluation frameworks found the five frameworks with social and organizational evaluation dimensions (the Kaplan framework mentioned above was included) lacked clarity or specificity of their evaluation criteria, or had inadequate focus to assess both individual and organizational aspects of evaluation [7]. In response the framework proposed by the reviewers included social, and organizational evaluation factors focused on the organization [24] not the broader environmental context (e.g., bureaucratic-political environment, populations) of a framework in the review [25].

Critical appraisers of HIT evaluation studies recognize the need for broader evaluation frameworks that include contextual considerations [23]. Health services research (HSR) approaches tend to be complementary to current information systems and informatics evaluation, drawing from a different range of disciplinary viewpoints and methods, and having a different focus. HSR has a long and productive track record in clinical [26] and programmatic evaluation studies. Its core principles are to focus on structure, process, and outcome of health services intervention [27]. The current study proposed looking to HSR as

### Table 1 Literature search terms and number of references retrieved

<table>
<thead>
<tr>
<th>Database</th>
<th>HIT terms</th>
<th>Evaluation terms</th>
<th>Number of references retrieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed</td>
<td>(Computer, record, documentation program, reminder, protocol, decision), in Major Mesh Heading (medical informatics, and in Minor Mesh Heading (e.g. computers, computer-assisted instruction, decision-support systems, hospital information systems, management information systems, medical record systems, micro-computers, radiology information systems, reminder systems, tele-medicine, attitude to computers)</td>
<td>Evaluation framework, measurement practice, evaluation model, evaluation research, evaluation studies, program evaluation/ methods, cognitive evaluation, usability testing, systems analysis, software evaluation, qualitative study, qualitative evaluation, focus groups, questionnaires, questionnaires, interviews, taxonomy, classification, balanced scorecard, outcome and process assessment (health care), cost-benefit analysis, health services research</td>
<td>2097</td>
</tr>
<tr>
<td>CINAHL+</td>
<td>Computerized physician order entry, CPOE, electronic order entry, health information system, clinical information systems, electronic health record, electronic medical record, computerized patient record</td>
<td>Evaluation framework, measurement practice, evaluation model, evaluation research, evaluation studies</td>
<td>545</td>
</tr>
<tr>
<td>EMBASE</td>
<td>Same as CINAHL+</td>
<td>Same as CINAHL+</td>
<td>6</td>
</tr>
<tr>
<td>AMED</td>
<td>Same as CINAHL+</td>
<td>Same as CINAHL+</td>
<td>0</td>
</tr>
<tr>
<td>Central</td>
<td>Same as CINAHL+</td>
<td>Same as CINAHL+</td>
<td>0</td>
</tr>
</tbody>
</table>

2. Objective

Because of the maturity of HSR evaluation, we expected to find a rich body of published HSR evaluations involving informatics, with a broader range of attributes about systems and contexts than informatics studies. This paper’s objective is to offer an HIT evaluation framework that uses HSR principles to address identified shortcomings in available HIT evaluation frameworks. In addressing this objec-
The study compared HSR and informatics literature related to HIT evaluation.

We based our methods on the principles of systematic evidence review, updating and extending with HSR terms, a list of evaluation criteria derived from a previous extensive literature review. We performed a search of informatics evaluation concepts across both HSR and informatics evaluation literatures. The domain of EHRs was selected to narrow the universe of studies because so much of informatics evaluation has been focused in this area and because of the importance of this domain in the national conversation. Studies that evaluated systems that were not specifically described as EHRs were excluded, such as studies of computerized provider order entry systems. We then compared HSR and informatics EHR evaluation literature, developed a framework of evaluation criteria for EHR; and used the framework to evaluate an EHR.

3. Methods

3.1 Literature Search

This systematic review updated/extended AdK’s prior exhaustive PubMed 1982–2002 review. Beyond AdK’s search strategy, the current study added terms to identify HSR articles about HIT evaluation terms (program evaluation/methods, health services research, outcome and process assessment (health care)). A standard systematic-review strategy was used with concept induction. Our method involved several high-level steps. First, reference databases were chosen to locate studies – CINAHL+, EMBASE, Cochrane Central Register of Controlled Trials (Central), and AMED – following the methodology of a model informatics evaluation article in Cochrane.

Second, a retrieval strategy used search terms (Table 1) for PubMed 2003–2007, and four other databases 1992–2007. For each database searched, all HIT terms were combined with OR, all evaluation terms were combined with OR, and HIT and Evaluation results were combined with AND.

Third, inclusion/exclusion criteria were established to reduce the resulting reference list. HIT inclusion/exclusion terms for informatics and evaluation study were identical to those in AdK’s review article. An HSR inclusion term for evaluation study was not found in HSR literature; as confirmed by an HSR evaluation expert knowledgeable in informatics (JW, acknowledgments). Instead, a working definition of evaluation from HSR was used: “systematic application of social research procedures for the purpose of making informed decisions”.

Fig. 1 Systematic review and elicitation of evaluation criteria; categorizing HSR and informatics studies
for assessing the conceptualization, design, implementation, and utility of social intervention programs” ([35] p 16) This definition is consistent with AdK’s terms for evaluation study.

Fourth, the lead researcher (PS) identified studies for exclusion that did not meet the study’s criteria for evaluation, HIT, or evaluation of HIT by checking title and abstract. The domain of electronic health records (EHR) was selected to narrow the universe of studies retrieved, as so much of informatics evaluation has been focused in this area and because of the importance of this domain in the national conversation. A standard EHR definition was applied: a digital repository of securely stored and exchanged patient data, accessible by multiple users to support quality and efficiently integrated health care [36] (which excluded systems such as computerized provider order entry-CPOE). The process to identify studies to be excluded achieved acceptable (80%) concordance with the second reviewer (PC). The second reviewer was an informatics student with an international public health background. Discrepancies were resolved by discussion.

Lastly, the contribution of HSR informatics evaluation literature was assessed. Articles were grouped as informatics or HSR by differentiating the EHR evaluation focus using AdK’s informatics evaluation studies criteria: “systematic, empirical assessment of a component of a health information system” and not an assessment of the impact on the health system or the patient’s health outcome ([31] p 45). To categorize articles as HSR, the primary condition was that a study assess EHR impact on the aspects of quality identified by Donabedian: structure, process, or outcome of the health system or the patient health outcome [37]. Journals were classified as HSR, informatics, or other (e.g., medical) by referring to the journal’s name, description of content, scope and aims, and ranking category from its web site.

### 3.3 Framework Formulation

The new framework was built on AdK’s multiaxial classification system that

#### Table 2 Evaluation framework and stakeholders

<table>
<thead>
<tr>
<th>Factor Level</th>
<th>Individual</th>
<th>Group</th>
<th>Organizational Systematic</th>
<th>Environmental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder</td>
<td>Patients</td>
<td>Clinicians</td>
<td>IT Department</td>
<td>Administration</td>
</tr>
</tbody>
</table>

#### Table 2 Evaluation framework and stakeholders

<table>
<thead>
<tr>
<th>Framework component</th>
<th>Individual</th>
<th>Group</th>
<th>Organizational Systematic</th>
<th>Environmental</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRUCTURAL QUALITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardware and technical quality</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Software quality, i.e. usability of the software</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Organizational support/capacity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Functionality</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>QUALITY OF INFORMATION LOGISTICS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completeness or correctness of data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Costs of information processing</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>User satisfaction</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patient concerns about security, privacy or confidentiality</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patient satisfaction, attitudes, perception toward HIT</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Diffusion</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

#### Effects on quality of processes

| Efficiency of work processes | X | X | X | X |
| Appropriatezza of patient care | X | X | X | X |
| Organizational or social quality | X | X | X | X |
| HIT selection/development, implementation and training | X | X | X | X |

#### Unintended consequences/benefits

| Diffusion | X | X | X | X |

#### Barriers or facilitators to adoption

| X | X | X |

#### Effects on outcome quality of care

| Morbidity, mortality, quality of life | X | X | X |
| Costs of patient care | X | X | X |
| Patient-related knowledge | X | X | X |

© Schattauer 2012 Methods Inf Med 2/2012
grouped evaluation criteria into four categories: Structural Quality; Quality of Information Logistics; Effects On Quality Of Processes; and Effects On Outcome Quality Of Care [31]. Transforming AdK’s structured collection of criteria into a framework entailed developing a multi-dimensional construct with interrelated and interdependent components. Each AdK evaluation criterion was changed from one-dimensional list to a two-dimensional construct, evaluation concept, which contained one or more evaluation components. Components were organized within concepts. The dimension, stakeholders, was added to include perspectives of providers and consumers of health care, which is consistent with HSR approaches that emphasize the variety of disciplinary perspectives [27]. Stakeholder types and their relationships to components were elicited from a convenience sample of informatics faculty. Table 2 provides a grid indicating the stakeholders.

Evaluation criterion from the above attribution-identification process was compared to AdK’s evaluation concepts/components to determine if the criterion was new to AdK and in which evaluation concept the new criterion fit. The structure/process/outcome framework provided a context for assessing fit. New criterion that fit into one of four AdK concepts was added to that concept. Criterion that incorporated and extended an existing component at the same contextual level and with the same stakeholder perspective replaced the existing component. Criterion that did not fit in a concept, yet provided more information about the components, such as a more granular description or a higher-level view, or included other stakeholder’s perspectives (e.g., patient), was added as a component to extend the concept. Criterion with a focus that was broader or at a different level than the concept’s (e.g., the criterion had an organizational focus; the concept had a clinician focus) was added as a new concept.

3.4 EHR Evaluation with Framework

Employing the evidence-based HIT evaluation framework in an evaluation of an electronic health record demonstrated the proof of concept of the framework. The framework was applied to an evaluation of a vanguard EHR in a nurse-managed community-based geriatric day care setting. This study evaluated the impact of an electronic health record (EHR) on clinician satisfaction with the care process. The EHR was implemented to maintain the site’s quality of care with increasing numbers of patients. Study subjects included 39 clinicians on the multi-disciplinary team caring for the elders. Data was collected at 11 and 17 months after EHR implementation. Quantitative data included: a clinician satisfaction survey and audit logs of EHR usage. Qualitative data was generated from observation of EHR use and interviews with clinicians and recorded in field notes. Institutional Review Boards approved the study.

Two methods were used to assess the framework: survey development and administration; and qualitative analysis of observations and followup interviews. The survey was a proof of concept of the utility of the framework. The framework provided the topics for the clinician satisfaction survey developed for this evaluation. Of the 20 evaluation components, all but three (i.e., Functionality, Costs of Information Processing, and Diffusion) were selected for the survey. At least one survey question, selected from existing health information technology satisfaction surveys, was matched to each framework evaluation component selected for the survey. The survey was administered at two points in time at two geriatric day care facilities, a site with an EHR and a paper-based site, to ascertain whether the survey could differentiate between sites. The survey was entitled EHRNS: EHR Nurse Satisfaction survey and greater detail about its development and validation is available [39]. The second method assessed the completeness of the framework. Field notes of observations and follow-up interviews were qualitatively analyzed seeking themes. Resulting themes were compared to framework evaluation components to assess the completeness of the framework. Further detail is available [40].

4. Results

We present the results of the literature review, the framework, and the framework evaluation. Literature review findings address the study’s objective to compare HSR and informatics studies related to HIT evaluation. Presenting the framework and the framework evaluation address the objective of offering an HIT evaluation framework that incorporates HSR principles.

4.1 Literature Search

Results were: 2097 PubMed references; 545 CINAHL+; 6 EMBASE; 0 Amed and Central. The resulting 2608 unique references were added to AdK’s yield of 1035 papers from 15,500 papers [31]. Applying inclusion/exclusion terms for informatics and evaluation studies to the title or abstract retained 353 articles (13%). Including only articles about EHR that met HIS and/or evaluation criteria retained 128 articles. Informatics and other journals had an even distribution of HSR and informatics articles; HSR journals had almost all HSR studies. Slightly more HSR (52%) than informatics (48%) articles were retained.

4.2 Elicitation of Evaluation Criteria

Seven evaluation criteria new to AdK were elicited from retained articles: together they occurred 28 times in total and none reached saturation. Informatics journals produced six of the seven new criteria, three of which did not appear in HSR journals: Patient Privacy; EHR Selection/Development/Training; Unintended Consequences. Five of the seven new criteria were elicited from HSR journals, two of which did not appear in informatics journals: Barriers/Facilitators to Adoption; Functionality. Other journals contributed five of the seven new criteria: all appeared in either HSR or informatics journals. All the retained articles (128) were read, as saturation was not reached for any evaluation criteria. HSR journals contributed 12% of studies and were 15 times more likely to contribute new criteria (15 (95% CI
2.8–84)) than informatics journals which contributed 53% of studies. (Other journals contributed 35% of studies.) HSR studies were 35% more likely to contribute new criteria (0.35 (0.14–0.86)) than informatics studies.

4.3 Framework Formulation

Each AdK evaluation criterion [31] became an evaluation concept which contained one or more evaluation components as shown in Table 3. We distinguish components from concepts by using SMALL CAPITALS. We also broadened concepts to apply to a range of stakeholders (e.g., patients).

Seven new evaluation criteria from the attribution-identification process were added to the AdK classification. Most new criteria were organized along AdK’s Structure/Logistics/Process/Outcome axes, but differ in content from AdK’s evaluation criteria. The first new component, FUNCTIONALITY, occurred in two reviewed studies, and was added to the AdK concept, Structural Quality. Functionality, an aspect of software, is consistent with Donabedian’s description of structure which includes tools and resources [37].

The criterion, DIFFUSION, describes HIT uptake within the organization, both in terms of degree to which clinical functionality is in use – how it is used (depth), and number of people using the function (broadth) [41]. Diffusion is a broader view of USAGE PATTERNS and replaces USAGE PATTERNS in the concept, Quality Of Information Logistics. The criterion, Diffusion, occurred in five studies. Two new components, TRAINING, occurring in seven new studies, and was added to the AdK concept, Structural Quality. Functionality, an aspect of software, is consistent with Donabedian’s description of structure which includes tools and resources [37].

The sixth new criterion, EHR SELECTION OR DEVELOPMENT, IMPLEMENTATION AND TRAINING occurred in seven studies and was added to the concept Effects On Quality Of Processes. The new component’s focus is the process of implementing new software in the organization, which extends the concept’s prior patient-care focus. Both are consistent with Donabedian’s Process: activities that constitute health care [37].

Table 3 Evaluation concepts and components from Ammenwerth and DeKeizer 2005

<table>
<thead>
<tr>
<th>EVALUATION CONCEPT</th>
<th>EVALUATION COMPONENT</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural quality</td>
<td>HARDWARE AND TECHNICAL QUALITY</td>
<td>System availability: access to computers, or sufficient bandwidth</td>
</tr>
<tr>
<td>Structural quality</td>
<td>SOFTWARE QUALITY</td>
<td>System usability: user-friendly i.e., cognitive aspects of the interaction of the user and HIT; general computer knowledge or user acceptance</td>
</tr>
<tr>
<td>Structural quality</td>
<td>ORGANIZATIONAL SUPPORT CAPACITY</td>
<td>Available resources; organizational and social context, such as collaboration, culture, and championing</td>
</tr>
<tr>
<td>Quality of information logistics</td>
<td>COMPLETENESS/ CORRECTNESS OF DATA</td>
<td>Data quality</td>
</tr>
<tr>
<td>Quality of information logistics</td>
<td>COSTS OF INFORMATION PROCESSING</td>
<td>Direct costs of purchased systems as opposed to costs of internally developed systems</td>
</tr>
<tr>
<td>Quality of information logistics</td>
<td>USER SATISFACTION/ ACCEPTANCE</td>
<td>Satisfaction with EHR</td>
</tr>
<tr>
<td>Quality of information logistics</td>
<td>USAGE PATTERNS</td>
<td>Which user used what functionality, i.e., how often a component was used</td>
</tr>
<tr>
<td>Effects on quality of processes</td>
<td>EFFICIENCY OF WORK PROCESS</td>
<td>Time the staff needs for a certain task (e.g. time spent entering and retrieving information from patient records)</td>
</tr>
<tr>
<td>Effects on quality of processes</td>
<td>APPROPRIATENESS OF PATIENT CARE</td>
<td>&quot;Medical efficiency&quot; such as adherence to protocols, a memory jog generated by presentation of information (i.e., seeing the result of a previous blood sugar test reminds the nurse to perform another test at the appropriate interval; or in a decision support system: adherence to clinical guidelines, or response to advice about possible medication errors)</td>
</tr>
<tr>
<td>Effects on quality of processes</td>
<td>ORGANIZATION/ SOCIAL QUALITY</td>
<td>Cooperation or communication among the care team, and includes inter-team communication</td>
</tr>
<tr>
<td>Effects on outcome quality of care</td>
<td>FUNCTIONAL OUTCOME</td>
<td></td>
</tr>
<tr>
<td>Effects on outcome quality of care</td>
<td>COSTS OF PATIENT CARE</td>
<td>Direct costs</td>
</tr>
<tr>
<td>Effects on outcome quality of care</td>
<td>PATIENT RELATED KNOWLEDGE</td>
<td>Patient’s knowledge of health condition</td>
</tr>
</tbody>
</table>
**Unintended Consequences/Benefits**, the seventh new criterion, occurred in two studies, and is the result due to lack of purposeful causation or action [42], such as benefits or happy surprises, or adverse consequences. **Unintended Consequences/Benefits** may touch on multiple concepts (e.g., Structural Quality (Organizational Support/Capacity); Effects on Outcome Quality of Care (Patient Outcome)); with a focus broader than a single concept, a new concept was created.

Transforming this structured collection of criteria into a framework entailed organizing evaluation concepts in relation to system use with logical links among the interrelated and interdependent concepts. The resulting Health Information Technology Research-based Evaluation Framework (HITREF) is shown in Figure 2 with new evaluation concepts/components indicated.

### 4.4 EHR Evaluation with Framework

The evaluation of the HITREF in the geriatric day care site with an EHR and the paper-base site demonstrated the utility of the HITREF as the survey was able to discern a change in clinician satisfaction at both the clinician level at each site and at the organizational level between sites.

Overall, clinicians \((n = 32)\) were satisfied with the impact of the EHR on the clinical process. Clinicians remained satisfied with the EHR’s impact on clinical process at Time 2 (17 months post EHR implementation), but were less satisfied compared to Time 1 (11 months post-EHR implementation). When clinicians’ survey responses were matched, 3 items had statistically significant differences between survey administrations among the 21 Likert scale items. **Sufficient Organizational Resources** (Item 20) responses indicated clinicians were satisfied and had decreased satisfaction \((p = 0.03)\). **Worth the Time and Effort** (Item 13) responses suggested clinicians were satisfied and had decreased satisfaction \((p = 0.03)\). **Clinician Involvement in HIT Selection, Implementation, Training** (Item 17) responses indicated clinicians were dissatisfied and had were less dissatisfied \((p = 0.03)\). Additionally, the survey instrument was able to discern a difference between the EHR site and the paper-based site \((n = 13)\) for 20 of the 21 items with Likert scale responses.

Assessing the completeness of the HITREF by identifying new themes that emerged from qualitative analysis of field notes of observations and interviews suggested that no new themes emerged and the HITREF was complete. Five major themes (i.e., occurred in both observations and interviews) and five minor themes (i.e., occurred in observations or interviews) emerged which mapped to nine HITREF components. Major themes included **Functionality**, a component new to AdK, and purposefully omitted from the EHRNS survey to be elicited from a different data source (as was Diffusion). The remaining major themes were also survey items: 1) system availability (Hardware and Technical Quality); 2) usability (Software Quality); 3) Completeness Or Correctness Of Data; 4) Efficiency Of Work Processes. Minor themes were also survey items: 1) system problems (another measure of Hardware and Technical Quality); 2) support (Organizational Support/Capacity); 3) Overall Satisfaction (User Satisfaction); 4) Team Communication (Organizational/Social Quality); 5) Clinician Involvement (HIT Selection/Development/Implementation/Training).
No theme emerged which did not map to a framework component.

5. Discussion

This research produced HITREF, the Health Information Technology Research-based Evaluation Framework, which is firmly grounded in research evidence and integrates principles of informatics and HSR evaluations. HITREF should serve as a conceptual tool for framing evaluations studies in assessing EHR-based implementations in organizational, systematic, and environmental contexts.

The commitment to evidence is a strength of this framework. HITREF was developed using a robust, reproducible, and productive methodology. An extensive literature review was conducted across five reference databases using replicable retrieval and inclusion strategies. This method was productive along two dimensions. First, the yield was 13%, twice AdK’s yield of 7%, 1053 papers from 15,500 [43]. Second, the review yielded 128 EHR evaluation studies which Ammenwerth added to the HIT evaluation online database [43]. The inclusion of HSR journals and studies was also productive. As hypothesized HSR journals and studies produced disproportionately more criteria new to AdK. Two of these criteria surprisingly did not occur in informatics journals when the study domain was limited to EHR (i.e., Functionality; Barriers/Facilitators to Adoption) although these criteria occur in the CPOE domain [41].

Our broadened investigation produced seven new evaluation criteria which addressed HIT evaluation framework shortcomings. Three components expanded the organizational context: Diffusion; EHR Selection Or Development, Implementation And Training; Unintended Consequences/Benefits. Also Barriers/Facilitators to Adoption added systemic and environmental (political) contexts. Lastly, three components added the patient perspective: Patient Privacy Concerns; Patient Satisfaction with EHR; Patient Satisfaction with Care. None of these new evaluation criteria reached saturation: broadening the inclusion of studies beyond the EHR domain may identify more occurrences of the new criteria or may identify more new evaluation criteria. The lack of saturation may also indicate that the new concepts are just beginning to appear in the literature reflecting recent changes in evaluation practice since AdK’s review, such as focus on unintended consequences [42], barriers or facilitators to adoption [44], and patient perspective [45].

Being evidence-based, this study may be limited in the criteria covered in evaluation studies of EHRs in contrast to papers discussing evaluation and evaluation methodology in the abstract. In excluding the latter, evaluation criteria available in these sources were not available to this study. Investigating this concern, juxtaposition of HITREF and a recent theory-derived framework, HOT-fit [24] indicated that evaluation criteria in the theory-derived framework were included in the evidence-based framework, and the theory-derived framework did not include the three new evaluation components related to patient perspective from the evidence-based framework.

The 20 resulting criteria were organized into the HITREF’s six evaluation concepts and their constituent components by employing AdK’s classification system to group criteria and Donabedian’s structure-process-outcome model to assess fit. The organization of new criteria within evaluation concepts or as new evaluation concepts followed the authors’ predetermined logic. These decisions may reflect current sensibilities. For example, the authors added Unintended Consequences/Benefits as an evaluation concept, which may reflect a current focus in the literature on unintended consequences. Other evaluators may classify Unintended Consequences/Benefits as an attribute of an EHR impact on outcome (e.g., Patient Outcome of Care, Patient Satisfaction, Efficiency Of Work Processes).

The HITREF was successfully used to guide clinician satisfaction survey development, and provide themes for multi-dimensional analysis across diverse data sources and interpretation of findings. Although the HITREF was drawn from a literature focused on physician HIT use in hospitals [31], in a nurse-managed community health care setting the HITREF was used productively and was a complete evaluation framework which included all themes that emerged from the evaluation. The HITREF, drawn from the HSR and HIT literatures, was intended to be a more complete HIT evaluation framework as compared to existing frameworks adapted from MIS literature or that have a specific HIT focus (e.g., system usage behavior). The HITREF, as a more complete framework, is relatively large and evaluators may choose to select evaluation components related to the questions being asked [46] or that reflect the evaluation goals of the evaluation site [47] and the researcher. For example, the initial negotiation establishing the scope and objectives of the evaluation should result in a clear, mutual understanding of the sponsor’s and the researcher’s evaluation goals, as well as the politically feasible questions [46]. The resulting questions would be matched to the HITREF evaluation components for inclusion in the evaluation. The HITREF is designed to be used to evaluate a broad range of HIT, reflecting the many HIT types which were evaluated in the underlying literature review. Although based on the EHR literature (because of the large number of articles published), we hope that the HITREF could be used to evaluate a wider range of HIT. The use of an evidence-based, comprehensive framework could make the resulting research findings more comparable and therefore more useful. Assessing the HITREF in diverse HIT settings (e.g., hospital, ambulatory, community-based, public health) and with a focus on various users (e.g., physician, nurse, interdisciplinary team) is an area of future research. Another research opportunity is to increase knowledge about the relationships among the evaluation components by using the HITREF in evaluations with sufficient sample size. The uncertainty of these relationships is a limitation which may impact evaluators’ immediate use of the framework.

6. Conclusion

The purpose of a framework is to ensure that a study is complete in the sense of including all the concerns that should be con-
Acknowledgments

NLM Informatics Training Grant T15LM07452, Robert Wood Johnson Foundation Public Health Informatics Training Grant; The Johns Hopkins School of Medicine; J. William Fulbright Fellowship for support. Jonathan Weiner for HSR consulting; Claire Twose for search strategies.

References


