A major challenge in the evidence-based design (EBD) practice model has been assessing evidence. Despite some individual variations in the way the design community conceives of the process, there is one fundamental underlying theme, as reflected in the EBD definition proposed in the Center for Health Design Web site (Center for Health Design, 2010), that of using “credible” research. What factors contribute to the credibility of evidence, and what are ways to evaluate evidence in practice? In this article, the term evidence is used instead of research. As discussed later, although findings from scientific research are considered to be one type of evidence (a crucial one), other sources of evidence exist and are currently being used in practice. The generation of such evidence should follow methods that enhance the credibility of findings.

Irrespective of its source, it is obvious that the more credible a piece of evidence, the more confidence one has in the findings for application in building projects. Although individual scientific
expertise to judge the credibility of evidence exists, currently there is no commonly accepted and understood mechanism for assessment that stakeholders in the larger community can use. Consequently, since the inception of the EBD paradigm, the challenge of assessing evidence has remained one of the major impediments to its broader adoption and constitutes a source of criticism (see, for instance, American Society of Healthcare Engineers, 2008; Stankos & Schwarz, 2007).

Several factors contribute to the difficulty associated with developing an assessment framework. The first problem is that a framework of this kind does not exist in the field of healthcare design; hence, it needs to be developed from scratch. Second, healthcare design draws on knowledge from a multitude of disciplines with varying approaches to knowledge generation. There exists the potential for divergent thinking among experts pertaining to the robustness of a particular study.

The central role of designers in the EBD model—professionals who are not trained to consume research literature—makes the development of an assessment protocol particularly important. The fact that all evidence does not and should not enjoy the same level of confidence could lead to its inappropriate use (perhaps, in some contexts, with negative consequences) if a simple, clear, and meaningful system of assessment is not developed for all stakeholders in the EBD process. The systematic assessment of healthcare design evidence could also identify knowledge gaps that should be the focus of future research studies.

Because EBD was inspired by evidence-based medicine (EBM), this article borrows a number of concepts and frameworks from the EBM and evidence-based practice (EBP) literature. With a substantial body of work and greater level of maturity in comparison to EBD, the medical disciplines have much to offer in the development of an evidence evaluation framework to support EBD. This paper draws from a small number of authoritative studies from the EBM/EBP literatures, while respecting and recognizing the differences between medical and design practice, as appropriate. The proposed framework attempts to appeal to the healthcare design research audience while maintaining simplicity and comprehensibility for stakeholders not trained in research.

Assessing and Utilizing Evidence

There are two distinct (and equally important) aspects to integrating evidence in design decision making; the first is the evaluation of a piece of evidence to assess its scientific merit. The second is the actual utilization of the evidence in design decision making. The former focuses on the scientific robustness of a piece of evidence; the latter pertains to the applicability of a piece or body of evidence to a specific context. One of the main reasons for separating assessment from utilization is the fact that the scientific merit of a piece of evidence is not synonymous with its applicability in a specific context.

A case in point is the experiment with an acuity-adaptable unit at Methodist Hospital, Clarian Health Partners (Hendrich, Fay, & Sorrells, 2004). Although the post-move data showed a substantial reduction in patient transfers, medication errors,
Because assessing the effectiveness, appropriateness, and feasibility of an intervention involves different objectives (the questions addressed are different), other forms of study design may also prove informative and more suitable for evaluating appropriateness and feasibility.

This distinction is not unique to the design industry and has been promoted in the medical disciplines (Evans, 2003; Stetler et al., 1998). Even though the medical disciplines initially started off with a single pyramid for grading evidence, subsequent literature has encouraged the separation of the two domains. Articulating a model for evidence-based nursing practice, Stetler et al. (1998) propose separate steps for reviewing methodological and utilization factors. Evans proposes three dimensions for assessment: effectiveness, appropriateness, and feasibility. These concepts have meaningful implications in the context of healthcare environment design, which is addressed later in this article.

What Is Evidence?

A critical notion underscored by Evans (2003) in his proposed hierarchy of evidence is the necessity to revisit the practice of seeking evidence originating mainly from classic randomized controlled trials (RCTs), or experimental studies. Because assessing the effectiveness, appropriateness, and feasibility of an intervention involves different objectives (the questions addressed are different), other forms of study design may also prove informative and more suitable for evaluating appropriateness and feasibility. Similarly, Miller and Jones-Harris (2005) propose an alternative evidence-based hierarchy that is related to Evans’. Their alternative framework suggests that the suitability of a study methodology is linked to the type of question being asked. They outline eight categories of questions that typically are addressed in clinical practice. In their hierarchy, appropriate study designs change as the nature of the clinical question changes, thereby expanding
the evidence base to study typologies (including a wide range of qualitative study designs) that were not considered evidence (or high-quality evidence) in previous models of EBM owing to the prior dominant emphasis on effectiveness.

The definition of evidence has been expanded further in the medical disciplines. Stetler and Caramanica (2007) propose that data generated from activities that technically are not scientific research can also provide evidentiary support (provided steps are taken to ensure their robustness). Distinguishing between external evidence (published research and nationally developed guidelines) and internal evidence (quality initiatives, implementation projects, program evaluations conducted within an organization, and affirmed experiences or group consensus), they propose that data from the latter sources qualify as evidence to support decision making in nursing practice.

This expanded definition of evidence is relevant to healthcare design decision making in many ways. First, the evidence base in healthcare design is not yet comprehensive enough to support all decision-making contexts. This is also true for evidence-based nursing practice (Stetler & Caramanica, 2007; Stetler et al., 1998). Even in the domains where substantial bodies of literature exist, the possibility of a lack of evidence to inform a specific clinical question is real (Haynes, 2006). Second, the nature of the question being addressed during the design process should dictate the type of evidence best suited to support decision making. The wider the evidence base, the better the breadth and depth of applicable evidence.

What are the unique contexts of healthcare design decision making? What questions are typically addressed in healthcare design? Can the EBM/EBP model be applied to healthcare design? What are examples of internal evidence in EBD? These and other questions are examined in the following sections.

Assessing Evidence
In general, the intent of the healthcare EBD process is to contribute to process and organizational outcomes through the optimization of the physical environment. This goal is not new; it existed even before the advent of EBD. The key difference is the addition of the use of credible evidence in optimizing the physical environment. Because the performance of processes and organizations is typically assessed using quantitative data, quantitative evidence in the realm of physical design is more amenable to performance estimation processes. This does not imply that qualitative evidence is of little use in the EBD process. On the contrary, it constitutes a rich source of evidentiary support for a number of questions during the facility procurement process, which is explored later.

The use of quantitative data facilitates EBD goals in another crucial way. Available study designs and analytical methods using quantitative data offer the potential to address one extremely important question, whether observed outcomes/changes are the effect of the physical design or
other factors actually are responsible for the effect (alternative explanations). This is also known as the internal validity of a study. Because estimating the impact of physical design on outcomes of interest constitutes the predominant objective of healthcare EBD, the ability to eliminate alternative explanations is vital, in general, to help determine the scientific strength of a piece of evidence. This is analogous to the evidence hierarchy in EBM/EBP, where study designs that are better able to eliminate alternative explanations are assigned a higher level on the evidence ladder.

Experiments in Healthcare Design Research
Experimental design constitutes the best method of eliminating alternative explanations. There are two essential components of experimental design that ensure the highest level of internal validity: (a) the presence of two or more groups for comparison, called the control or comparison groups, and (b) random assignment.

Designing and conducting a true experiment in healthcare design research—although challenging owing to ethical, logistical, and other reasons—is not impossible. One example is a study comparing a standard consultation room with a new consulting room design concept that enhanced patients’ visual access to the computer screen and lessened psychological separation from the physician (Almquist et al., 2009). To examine whether the new design concept had any impact on six domains of interest, the team redesigned two consultation rooms to reflect the standard Mayo Clinic model and the experimental model. Patients were randomly allocated to the two room variants, and data on outcome variables were collected using questionnaires.

More commonly, researchers have attempted to find naturally occurring variations between settings to conduct studies that offer comparable levels of internal validity. Natural experiments are best suited to situations where attributes of the environment other than the environmental element of interest, processes, and users are identical between settings. Ulrich’s (1984) study on exterior view content and its impact on surgical patients is an example of this. Owing to the random assignment of patients based on bed availability, the assignment of patients to two sets of rooms—one with view of a small stand of deciduous trees and the other of a brown brick wall—could be considered random. Operational and organizational attributes were identical for the patients in both groups. The rooms were also identical. As reported, all rooms were double occupancy and largely similar in dimensions, window size, bed arrangement, furniture, and so forth. Furthermore, patients in the two groups were matched based on such attributes as age, sex, smoking habits, weight, and so on.

Another example involved a study to examine the differential influence of inboard versus outboard bathroom locations on clinicians and patients. The researchers selected a unit in which half of the patient rooms had outboard bathrooms and half had inboard bathrooms (Pati, Harvey, & Ragan, 2009). During the period of data collection, nurses were randomly assigned to patient rooms with either outboard bathrooms or inboard bath-
rooms, but not both. Because every other attribute remained identical, the study design ensured a high level of internal validity.

Many clinical studies also provide credible information on the physical environment when it is considered to be a moderator (interacting variable) in clinical care. For instance, in a study that aimed to examine the influence of sunlight on patients with a first episode of myocardial infarction (Beauchemin & Hays, 1998), the researchers selected two sets of patient rooms, one bright and southerly, the other dark and northerly. Because patient bed assignment is subject to bed availability, the process of assigning patients to a northerly or southerly room could be considered naturally random. To further improve the uniformity (or equivalence) of the two groups, the team excluded patients with several characteristics, such as those with major surgery, those not directly admitted to the unit, and so forth. Patients in the bright southerly rooms were considered in this example as the experimental or treatment group (presence of sunlight), and those in the dark northerly rooms were considered the control group (absence of sunlight).

Healthcare design studies also adopt computer simulation tools to design experiments. Although not all simulation studies examine causation, a typical attraction of simulation studies is the ability to design experiments in contexts where it may be cost prohibitive, time prohibitive, or impossible to design an experiment in real settings. One example is a bed-tower project in which the design team was unable to conclude whether medications and medication-related supplies in a central medication room represented a better solution than the provision of bedside medications in a nurse server inside each patient room. The performance dimensions of interest included staff walking distance and the time spent in medication-related activities.

With limited time available, a viable approach was to design an experiment, but in a simulated environment. MedModel is a process simulation tool that provides a powerful analytical technique for aiding decision making. Ethnographic observations and interviews were conducted to ascertain the factors that affect nursing staff efficiency, such as the daily average number of trips needed to the medication room and to nurse servers for medical and surgical patients, the average time to access and transport medication, and the average time to deliver medications to patients. Two floor layouts were used in the experiment, in which everything was identical except for the provision of medication. Nurses in the model were instructed to deliver medications typical during 12-hour shifts. Performance data generated from the exercise were compared to aid decision making (Pati & Barach, 2010).

Simulation studies can also involve analyses using principles of physics to predict or estimate the future state of building energy systems, acoustical environments, lighting conditions, and so forth. These studies are not identical to RCTs in the medical disciplines; however, they typically involve the counterpart of the control group (the base condition) and experimental condition(s).
Attributes of the physical environment are systematically manipulated and compared with the base condition in one or more performance measures. The researcher-controlled manipulation and the cause-effect nature of the study, along with comparison conditions, qualify these studies as experiments. Such studies can be valuable to healthcare design, especially when estimating potential future states involving new concepts.

A case in point is a study that focused on understanding how clean-room technologies might alleviate airborne contamination if adopted in hospital design. Airborne infections have been documented as a major source of hospital-acquired infection—one of the major concerns in healthcare delivery today. An important factor contributing to airborne infection is cross-contamination via air particulate dispersion as affected by ventilation system design. Clean-room technology (with membrane ceiling) has been used successfully in the technology and pharmaceutical industries to control airborne contamination. This study examined the performance of membrane ceiling technology in controlling air particulate dispersion in an exam room. It included both performance tests in a full-scale mock-up room and a simulation study of six different ventilation system designs using computational fluid dynamics (CFD) analysis (Pati et al., 2010).

A challenge posed by computer simulation methods is that the quality of outcomes from the study depends on the quality of input data. In the membrane ceiling study, preliminary outputs from the CFD analyses were validated using actual smoke tests in the mock-up exam room. In the preceding nurse server study, the use of data from ethnographic studies provides assurance of the accuracy of the estimate. Some form of validation of the outcomes of simulation in real-life environments enhances confidence in the findings.

**Quasi-Experiments in Healthcare Design Research**

A class of design with a lower level of assurance with regard to internal validity is *quasi-experimental design*. This class of studies is identical to experiments except that there are no random assignments. The lack of random assignment increases the potential for alternative explanations. This type of study is easier to implement in field research and hence more common. The elimination of alternative explanations depends on logical arguments rather than methodological techniques. For instance, in a study examining the influence of single-family neonatal intensive care unit (NICU) rooms on staff stress and satisfaction compared to traditional open-bay configurations, the researchers collected data from two hospitals' NICUs with open-bay and single-family room configurations (Shepley, Harris, & White, 2008). There was no random assignment of staff to the two conditions. To enhance internal validity, the researchers conducted comparative analyses of staff attributes to demonstrate that staff profiles were not significantly different in the two hospitals.

Many times, random assignment to two groups, such as groups of males and females or right-handed and left-handed staff, may be impossible. In such contexts, the best option available is the qua-
si-experiment. For instance, in a study to examine whether same-handed care environments (specifically, direction of approach) result in standardized behavior, an experimental setting was used to compare two groups of nurses, left-handed and right-handed. The physical configuration was manipulated as the nurses performed predefined tasks to examine whether the attributes of the physical configuration changed the subjects’ behavior (Pati, Cason, Harvey, & Evans, 2010).

A common form of study in the quasi-experimental category is the before-and-after study design. A number of examples of these studies exist in the Pebble Project (http://www.healthdesign.org/research/pebble/), a research program within The Center for Health Design. Before-and-after studies in healthcare design have typically involved collecting baseline data before moving into a new facility and collecting a second set of data on the same measures after moving into the new or renovated facility. These studies are weaker for examining causation, and they warrant logical arguments on a case-by-case basis to identify and eliminate alternative explanations. Because a large number of physical design elements (and also, perhaps, process attributes) change between the old and new facility, the causal links between elements of the physical environment and changes in outcome are less certain.

The attractiveness of the before-and-after design is the relative ease of implementation in healthcare design research, compared to a true experiment. In addition, before-and-after studies constitute a rich source for exploring and identifying new phenomena or relationships between the built environment and outcomes related to people and processes. They also serve well to examine the potential influence of new design interventions in healthcare settings, even when conclusive inferences on causality are impossible.

A before-and-after design is frequently used to conduct Lean studies to eliminate waste from healthcare processes. Process changes often involve physical environment changes as well. One such published study was conducted at the Virginia Mason Medical Center, Seattle (Nelson-Peterson & Leppa, 2007). In this study, nursing rounds and assessment processes were revised; supplies were kept at the point of care and restocked using a visual flag system; patient assignments were organized for work sequences to flow in an efficient, U-shaped format; and charting was performed at bedside. Outcomes from the Lean process implementation included fewer patient complications, increased staff time at bedside, and reduced walking distances.

Quantitative studies that do not attempt to examine causation are nonexperimental studies.

Nonexperiments in Healthcare Design Research

In general, quantitative studies that do not attempt to examine causation are nonexperimental studies. For instance, studies examining correlations are nonexperimental, although the existence
of correlation is necessary to establish causation. In environmental design there are many classes of post-occupancy evaluations (POEs) that use quantitative study designs that can be classified as nonexperimental studies. According to one definition, “post-occupancy evaluation is the process of evaluating buildings in a systematic and rigorous manner after they have been built and occupied for some time” (Preiser, Rabinowitz, & White, 1988, p. 3).

POEs in architecture have been conducted from several different perspectives. At one end of the scale are those interested in the performance of building systems and subsystems (envelope, roof, flooring, structure, heating, venting, and air conditioning, and so forth). At the other end of the scale are those whose interest lies in evaluating the functional performance of the built environment. Many studies fall somewhere between these two. Quantitative POEs that involve one-time post-occupancy data collection (the most commonly executed type) are nonexperimental in design.

Surveys that use quantitative data to understand/identify user needs at the beginning of a project and user satisfaction are nonexperiments. Space optimization studies such as those based on Serviceability Tools and Methods® (http://www.icf-cebe.com/tools/tools.html) are also essentially nonexperimental. An example of a quantitative study pertaining to space that is not causal in nature is a published study on healthcare design trends (Latimer, Gutknecht, & Hardesty, 2008). The authors analyzed data from their company archives to articulate a trend in space requirements and actual programmatic areas in hospitals over 2 decades.

Another frequently used nonexperimental technique in healthcare design is the environmental mock-up. Mock-ups are used to evaluate environments during various stages of the design and construction process. Typically they are not used to draw causal inferences. They are essentially one-off case studies that are designed to inform specific projects. Mock-up studies can involve both quantitative and qualitative data.

Qualitative Studies in Healthcare Design Research

In general, studies that do not use quantitative data are called qualitative studies. Qualitative data can include transcripts of interviews and focus groups, gaming, observation data, and so forth. Such studies typically do not examine causation. However, qualitative studies follow the same scientific process as quantitative studies. They offer the most appropriate methodology for many contexts, such as while exploring a new area of interest or when the richness of the data is crucial. Although such studies may not provide causal inferences, they help explore associations

A qualitative approach could be adopted in most situations other than studies with an objective to explain and predict.
between variables and potential patterns of relationships, and they could help propose new theoretical frameworks. In fact, a qualitative approach could be adopted in most situations other than studies with an objective to explain and predict. There are techniques for using qualitative approaches to explain and predict (see, for instance, Miles & Huberman, 1994); however, such approaches are not widely published in the domain of healthcare design research.

One context for the optimal use of the qualitative study design is when better understanding of complex concepts becomes critical to optimizing physical design. One such study focused on the concept of flexibility. Whereas designers approach healthcare design with their own understanding of the term (and the associated professional bias), there was little knowledge available on what flexibility meant to caregivers and the domains of design decision making that influenced operational flexibility. The best approach to examining the issue was qualitative. Semistructured interviews were conducted with stakeholders from six departments at six major acute care hospitals across the United States. Content analysis of the transcripts identified nine domains of design decision making that influence operational flexibility (Pati, Harvey, & Cason, 2008).

Qualitative studies that can serve as evidence need not always have a strong theoretical foundation. Examples of such studies in environmental design include POEs and mock-ups. One type of qualitative POE constitutes a quick walk-through in the early (or immediate pre-occupancy) phases of the life span of a building. They typically involve short periods of study, using small teams of stakeholders, and they result in text-based recommendations for the improvement of a facility. The recommendations may be for immediate, short-, or long-term measures, but they are generally focused on improving a single facility. Group discussions, touring interviews, photography, and expert observations are some of the tools used frequently. Such POEs are gradually increasing in frequency in healthcare design practice, although they are rarely published. Published examples of this type of POE in other settings include work by Fitzgerald and Tramposch (1998) and Watson (2003).

Similarly, studies to understand user needs can involve qualitative approaches, including one-on-one interviews, focus groups, gaming, observations, and so on. Mock-ups generating qualitative data are also considered qualitative studies, whose findings are useful in design decision making.

**Systematic Reviews in Healthcare Design Research**

Systematic reviews involve the review of multiple studies on a single question. The inclusion of multiple research studies provides higher credibility compared to a single study. Systematic reviews involving quantitative controlled studies are called *meta-analyses*, studies involving established statistical procedures that use the end product of multiple studies with related hypotheses as input for the analyses. Systematic reviews can be conducted on quantitative studies without controls as well as on qualitative studies. Stetler (2002)
makes a useful distinction between systematic statistical review and systematic interpretive review to appropriately describe the systematic review of controlled quantitative studies versus other quantitative and qualitative studies.

In essence, the context of systematic review in medical disciplines is different from healthcare design. Literature reviews in medical disciplines are targeted toward specific clinical practices and interventions. Such specificity is not the nature of design practice. The context of each design is unique, different from every other context in meaningful and significant ways. This focus on specific clinical interventions has enabled medical disciplines to generate bodies of research on narrowly defined questions that can be subject to systematic review. This is not true in healthcare design, and systematic statistical reviews (meta-analyses) on healthcare design questions are virtually nonexistent.

Literature reviews on broad topics of interest in healthcare design are available and increasing in number. Examples include an extensive review of the research literature examining the impact of physical design on healthcare outcomes, which actually focused on multiple topics of broad interest (Ulrich et al., 2008; Ulrich, Zimring, Quan, & Joseph, 2004). Reviews on narrower topics are also available, such as one on lighting produced by The Center for Health Design (Joseph, 2006). However, such reviews serve more as an (extremely informative) assessment of the current state of evidence involving both quantitative and qualitative studies rather than integrative reviews (as suggested in EBP literature; Stetler et al., 1998) to assist in effective and cost-efficient design decision making on a targeted topic. The reason for this could be that, unlike in the medical disciplines where specific targeted interventions can be designed to address narrowly defined care, such situations do not yet exist in healthcare design practice.

Perhaps as environmental design, in general, and healthcare design, in particular, mature in the use of evidence to support design decision making, the generalizability of physical design interventions will be better understood, and integrative reviews supporting such decisions will emerge. In the current context, it is the author’s opinion that integrative reviews should be conducted for specific projects, separately, and general literature reviews should be used to understand the state of evidence and as a quick source of evidence for integrative reviews.

Levels of Evidence in Healthcare Design

The first system for the classification of EBD evidence, as in the case of EBM/EBP, constitutes the study design. This structure is based purely on the notion of internal validity. Among the hierarchies available in the EBM/EBP literature, one system developed by Stetler (2002) can be easily adopted for EBD with some qualifications. Stetler’s hierarchy includes eight levels of evidence based on quantitative and qualitative studies as well as nonscientific pursuits. Table 1 shows the hierarchy with sample studies from healthcare design.

Because examples of levels I, II, and V study types have yet to emerge in a clear fashion in healthcare design.
design literature, they are left as levels for future use. In the current state of EBD literature, levels III, IV, VI, VII, and VIII exhibit the greatest potential for use in practice.

For end users unacquainted with research design, simple questions can help determine the exact level of a specific piece of external evidence in healthcare design research. Ask whether the study is attempting to establish causation between a physical design concept/element and outcomes. If the answer is affirmative, the study is either a level III or level IV study. If the study involved random assignment (either researcher-manipulated or naturally occurring), it represents a level III study. If not, it represents a level IV study.

If a published study does not attempt to establish causation, it represents one of the remaining classes: levels VI, VII, or VIII. Level VIII is self-explanatory. The only confusion for the end user could be in deciding between levels VI and VII. Level VI evidence represents the body of studies that follow established steps of scientific research. Most studies in this category are based on a theoretical framework and proceed through a system-
atic scientific approach. Such studies include discussions on the generalizability of study findings.

The objectives of level VII evidence do not include generalizability to a larger population or to a theoretical framework. Their main purpose is to evaluate design concepts or built environments, the findings of which are intended for use in a specific facility design. When such studies are conducted in a systematic manner—“obtained in a manner that is replicable, observable, credible, verifiable, or basically supportable” (Stetler & Caramanica, 2007)—the findings from individual studies as well as from multiple studies can represent a highly credible source of evidence for design decision-making support.

Quality of Evidence
Although, in general, the research design adopted provides assurance regarding the degree of internal validity, when not properly addressed, quality issues can produce poor-quality research. Such issues are accounted for in evaluation frameworks in EBM and EBP (see, for instance, Harbour & Miller, 2001; Miller & Jones-Harris, 2005).

In general, quality evaluation in the medical disciplines proceeds through a series of methodological questions that are expected of high-quality research. Widely accepted evidence quality evaluation frameworks have been developed by the GRADE working group (see, for instance, Guyatt et al., 2008a; Guyatt et al., 2008b), the AGREE Collaboration (The AGREE Collaboration, 2001), and the University of Oxford, Center for Evidence Based Medicine (Phillips et al., 2009).

The (A–D) below each level of evidence in Table 1 represents the four grades a piece of evidence can be assigned to based on methodological issues in each tier of scientific strength.

The evaluation of scientific strength (Table 1) also includes adjustments required to determine the quality of a piece of evidence. The (A–D) below each level of evidence in Table 1 represents the four grades a piece of evidence can be assigned to based on methodological issues in each tier of scientific strength. It is expected that these adjustments are made by people knowledgeable of research quality issues and that the quality assessment is conducted in a peer-review format. It is expected that the grades are determined using a typical, academic-style grading system, where A represents excellent and D represents poor.

Table 2 outlines the issues that must be systematically examined by experts to determine the quality of a piece of evidence. Stetler’s quality evaluation framework offers a good list of quality issues (hence, it has been adapted for this suggested EBD framework) in scenarios involving a group of expert healthcare design researchers collaborating to evaluate evidence, or
a dedicated team that includes people trained as researchers in each individual healthcare project. The framework in Table 2 is meant to be further scrutinized for the field of healthcare design and changes made, if necessary, to reflect real-world differences between EBP and EBD.

**Table 2. Quality Assessment Framework for Individual Study Findings**

<table>
<thead>
<tr>
<th>Assessment Domain</th>
<th>Grading Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose and Overall Method</td>
<td>Was there a fit between the research question and each progressive piece of the methodology? Were there any obvious areas of potential bias in the data or its interpretation?</td>
</tr>
<tr>
<td>Sampling</td>
<td>Was the sample size too small to make credible comparisons between subgroups or assess a group’s characteristics? (For example, did they do a power analysis for an experimental design? Also, look for the final number of subjects in terms of those who actually completed the study.)</td>
</tr>
<tr>
<td></td>
<td>Did the sample match the research question?</td>
</tr>
<tr>
<td></td>
<td>For qualitative studies, were sufficient examples, experiences, or other information obtained?</td>
</tr>
<tr>
<td>Research Design</td>
<td>Did the design match the research question/hypothesis?</td>
</tr>
<tr>
<td></td>
<td>In an experimental type of design, were there any plausible alternative explanations for the findings, and was there a sufficient difference between the groups to effect a true difference in outcomes?</td>
</tr>
<tr>
<td></td>
<td>As applicable, was there an appropriate comparison or control group? Were relevant intervening factors controlled in some way (e.g., by random assignment, measurement, or exclusion of certain types of subjects)?</td>
</tr>
<tr>
<td></td>
<td>For qualitative research, was there a clear methodology, such as grounded theory?</td>
</tr>
<tr>
<td>Measurements</td>
<td>Was validity of the tools addressed and acceptable?</td>
</tr>
<tr>
<td></td>
<td>Was the reliability of the tools and the data collection process addressed and acceptable?</td>
</tr>
<tr>
<td></td>
<td>Were multiple measures used in the study (i.e., triangulation)?</td>
</tr>
<tr>
<td></td>
<td>For qualitative data, is there any way to confirm the researcher’s interpretations, e.g., did they use the subjects’ validation?</td>
</tr>
<tr>
<td>Analysis</td>
<td>Do the interpretations of the researcher go beyond what the data support?</td>
</tr>
<tr>
<td></td>
<td>Did the author differentiate statistical and practical significance?</td>
</tr>
<tr>
<td></td>
<td>For a qualitative study, can one logically follow the researcher’s data analysis?</td>
</tr>
<tr>
<td></td>
<td>Were meaningful, appropriate statistics used? Does it sound like the author is dredging for positive results?</td>
</tr>
</tbody>
</table>

*Source: Reprinted from Stetler, 2002, with permission from Dr. Stetler.*
Utilizing Evidence

As mentioned previously, utilizing a piece of evidence to support design decisions is different from assessing the scientific strength and quality of the evidence, although both can be done simultaneously. The steps recommended for a typical healthcare design EBD process have been articulated in the existing literature, specifically in the evidence-based design accreditation and certification (EDAC) study guide (Center for Health Design, 2009). This discussion specifically relates to the “Critically Interpret Relevant Evidence” section of EDAC Study Guide 3 (Center for Health Design, 2009).

Because the context of each healthcare project is unique, it is prudent to review available evidence within the context of its application. The examination of the applicability of evidence in a specific context requires a thorough evaluation of the similarities between the resultant sample of a study and the users of the facility being designed, the similarities between the cultural attributes of the setting in the evidence and the one being built, study limitations that may affect applicability in a specific setting, factors unique to the facility being designed, whether the application of a concept is culturally and financially feasible in the context in question, and so on. In other words, the cultural, financial, and legal boundaries within which a facility exists influence the applicability of evidence. Moreover, because EBD is not restricted to one nation or region, examination of the evidence-context fit constitutes a vital task, separate from the evaluation of scientific strength and the quality of evidence. Scientific strength and quality of evidence do not change across contexts, but appropriateness and feasibility do.

The EDAC Study Guide 3 outlines a process that involves the critical evaluation of individual studies, articles, and reports, once the information repository system has been developed in a specific project. The methodology and rigor associated with the concept of integrative review (IR) in EBP (Stetler et al., 1998) offer a potent tool for evaluating the evidence-context fit during critical reviews. IRs should not be confused with literature reviews, because these reviews also evaluate the appropriateness and feasibility of evidence in a particular context. If multiple pieces of evidence (quantitative and qualitative) converge to support a design decision or concept, it is better viewed through an IR process and from within the framework of cultural-legal-financial constraints.

The assessment of appropriateness and feasibility is best conducted jointly with team members from the client organization to understand and take into account the uniqueness of a context. Because time frames for evaluation and human resources vary among healthcare projects, it is impractical to develop an evaluation protocol that will fit all situations. It is anticipated that, as individual projects experiment with evaluating
the appropriateness and feasibility of evidence, a protocol and tools will be published to enable subsequent examination of whether a commonly applicable model is feasible.

The concepts of appropriateness and feasibility (Evans, 2003) have different meanings in healthcare and in clinical practice. From an EBD perspective, appropriateness is whether a piece or body of evidence is applicable and meaningful in a particular context. For instance, one could ask whether the acuity-adaptable room concept (Hendrich et al., 2004) is applicable to a given project considering the targeted patient population, census, size of the unit/hospital being designed, and so forth. Feasibility pertains to successful implementation. Feasibility questions include such challenges as successfully changing the culture to implement the acuity-adaptable model as well as the capital cost of implementing the concept, and operational cost once the facility is occupied.

Examining appropriateness and feasibility requires a deep understanding of the client organization and its users. The most meaningful sources of evidence for understanding the client and users may not necessarily be evidence of high scientific strength (Table 1). One may resort to lower-strength evidence, both internal and external (Stetler & Caramanica, 2007), such as POEs and user surveys, to facilitate better understanding of the context. Internal documents on patient demographics, needs analyses, and other data collected as standard operational protocol can also serve as useful evidence (Center for Health Design, 2009). Ensuring that the internal evidence—whether as standard operational protocol or project-specific data collection—is obtained with an emphasis on reliability and validity will provide high-quality evidence to support decision making.

Sample questions to help determine appropriateness for a client organization and its users include:
• What is the current state of organizational culture and operational philosophy?
• What is the desired future state of organizational culture and operational philosophy?
• What are the key organizational objectives and business drivers?
• What is the current user experience (for staff, patients, visitors)?
• What are the key challenges to current operations?
• What impediments does the facility design pose?

Sample questions pertaining to feasibility include:
• Can the patient care culture be changed successfully to enable the successful implementation of the physical design concept?
• How will the concept implementation impact capital and operational cost?

Although no protocol for IR in EBD is being proposed at this juncture, some attributes of the review process are important to the quality of the outcome of the process. Those characteristics should be considered a part of any IR in EBD. A peer review format should be adopted for evaluating studies; ideally, evaluation of evidence should not be a one-person job. A tabular format should
be used to document key information from the evidence systematically. A single table per study that can later be synthesized into tables that draw standardized information from multiple evidence tables should be considered.

The EBP model (Stetler et al., 1998) also envisages two separate tables for the presentation of key information from the IR of a study. One table captures methodological factors and the second captures utilization factors. Tables 3 and 4 show the main classes of information and details necessary for integrative reviews. Key to the review process is to objectively record facts and evidence rather than opinions and interpretations. Depending on the scope of the project, the volume of evidence being examined, the availability of resources, and the importance of the question, the review process in EBD can be conducted in two rounds: the first round, involving independent assessment, followed by a consensus round to eliminate biases. Multiple tables may be needed to best support the synthesis process for studies that involve multiple domains of findings.

**Surrogate Assessment**

In healthcare EBD, one important audience group is the designer end user, who is not trained to read and digest research publications. This fact constitutes the most challenging aspect of the EBD practice model, in the sense that one vital group of stakeholders is not conversant with the most critical aspect of EBD.

Furthermore, the medical disciplines have witnessed the emergence of core expert groups, resulting in the centralization of the process of evidence evaluation (although the option of conducting evaluations on a local scale has not been eliminated). Examples of such expert groups include the AGREE collaboration (http://www.agreecollaboration.org/), the GRADE working group (http://www.gradeworkinggroup.org/FAQ/index.htm), the University of Oxford Center for Evidence Based Medicine (http://www.cebm.net/index.aspx?o=1025), and the Cochrane Collaboration (http://www.cochrane.org/), among others. Such collaborations constitute reliable and authoritative sources for clinical practitioners in the field. Similar efforts have not surfaced in EBD practice. Consequently, assessment of evidence is largely a local effort in design practice, and a large burden is placed on design professionals to estimate the credibility of a piece or body of evidence.

The fact that only a small proportion of design firms has researchers on staff or access to someone with research expertise poses a major challenge. A scenario involving a group of expert healthcare design researchers collaborating to centralize the process of evidence evaluation could, perhaps, allow a greater degree of complexity to the assessment framework and the depth of evaluation.

Until the emergence of such a collaboration to evaluate evidence, and in the absence of personnel trained to evaluate the strength and quality of evidence, a simpler strategy is suggested here. This strategy is not intended to replace a thorough, formal evaluation of individual study findings; it is intended to serve as a surrogate
### Table 3. Classes of Information and Details Pertaining to Methodological Factors

<table>
<thead>
<tr>
<th>Methodological Information</th>
<th>Important/Informative Details To Include</th>
</tr>
</thead>
</table>
| Title, Year, Author, and Source | Information pertaining to credibility of author/s  
Country where study was conducted  
Important direct quotes from the publication or document that describe the data |
| Purpose of Study and Hypothesis or Study Question | Direct quotes from the publication |
| Measurements and Operational Definitions | Independent variables and operational definitions  
Dependent variables and operational definitions  
Intervening variables (mediators) and operational definitions  
Interacting variables (moderators) and operational definitions  
In an experimental or quasi-experimental study, clearly articulated physical design interventions being compared  
Tools used for measuring variables  
Brief note on reliability and validity of the tools  
Numerical anchors and possible range of scores for tools with scales  
Conceptual definitions of variables, if important |
| Sampling | Sampling method  
Sample size—both overall and subgroups  
Response rate |
| Design and Level of Evidence | Brief description of study design  
Theoretical framework, if used  
Scientific strength and quality of evidence (Table 1) |

Source: Adapted from Stetler et al., 1998.

### Table 4. Classes of Information and Details Pertaining to Utilization Factors

<table>
<thead>
<tr>
<th>Utilization Information</th>
<th>Important/Informative Details To Include</th>
</tr>
</thead>
</table>
| Findings | Practical significance of study findings (for a specific project)  
Both significant and nonsignificant findings  
Actual magnitude of change, if reported in the publication, and data relevant to show the nature of variation  
Relevant conditions important to the application of the concept |
| Fit | Sample: similarities between the attributes of the resultant study sample and the targeted users of the facility being designed  
Context: key similarities and differences between the study setting and the facility being designed (cultural, organizational, environmental, etc.)  
Implications: practical implications for the processes and people (patient, staff, visitors) if the concept is implemented (Estimate objective predictions on important outcomes measures if data are available) |
| Unknown Factors | Lack of information on any factor/s important to the context of application  
Study limitations that have implications for application in the context in question |
| Implications for Feasibility | Challenges specific to the context of application, such as cultural change |
| Cost-Benefit | Estimated magnitude of change in outcomes in the context of application  
Data on estimated capital cost and operational cost, if available |

Source: Adapted from Stetler et al., 1998.
estimate of the scientific strength of a piece of evidence. It can be viewed as a quick surrogate estimate for design practitioners who may not have the resources to approach a researcher every time they encounter a piece of evidence. Moreover, the need for an estimation of strength and quality does not always occur within the context of a well-resourced healthcare project.

The strategy proposed here is to associate a star (*) with every quality issue reported in an individual published study (not including literature reviews). A high-quality study is expected to address the quality areas outlined in Table 2. In turn, a high-quality publication of study findings is also expected to report on each of the quality issues. Although complete reporting on all quality issues does not necessarily guarantee high quality (and vice versa), an assumption is made here that publications that report on aspects of study quality are more transparent (and possibly of better quality because these issues were considered) than publications that do not report on quality issues. This is particularly important because the framework proposed in this paper includes information sources beyond traditional scientific publications, and quality issues may not be discussed in many such sources of evidence. Multiple sources of evidence are being used in current EBD practice, and, in many situations, estimating quality can be confusing.

Table 5 outlines the information to look for in an individual published study to apply the star-system surrogate quality estimator. The proposed quality evaluation starts with a five-star rating for an individual piece of evidence, but it loses a star for every aspect in Table 3 that is absent or not reported. Thus, by default a study is of five-star estimated quality (or excellent). Depending on the source of evidence or the quality issues reported in the document, an individual study finding can lose all stars and represent evidence of estimated poor quality. The quality estimation here is based purely on a binary (yes/no) scale, using a checklist to determine whether or not certain information is provided in a document.

Some illustrations of how to use this evaluation system will clarify the proposed interim framework. One example is the consultation room study described previously (Almquist et al., 2009). As an individual experimental study, it belongs to level III in scientific strength. As regards quality, it is published in a peer-reviewed journal (Health Environments Research & Design Journal). In addition, the published article includes all the information outlined in Table 3, except no blinding was involved. A patient–clinician interaction framework and patient-centered care model are introduced in the beginning. The objective—to understand the extent to which the experimental set-up affects patient–clinician interaction—is clearly articulated. The method of sampling the patient population and the use of power analysis to arrive at the minimum required 33 participants are explained. The methods, setting, participants, interventions, and data collection are clearly articulated. The six domains of interest are described. The practical significance of differences found in
Table 5. A Surrogate Quality Estimation Framework

<table>
<thead>
<tr>
<th>Estimation Factors</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of Evidence</td>
<td>Peer-reviewed journal, peer-reviewed conference proceedings, academic dissertation</td>
</tr>
<tr>
<td>Purpose and Overall Method</td>
<td>A theory or framework is introduced within which the study is conducted</td>
</tr>
<tr>
<td>Purpose and Overall Method</td>
<td>The research question and objectives are explained clearly</td>
</tr>
<tr>
<td>Sampling</td>
<td>The document reports how the sample size was determined, for both quantitative and qualitative studies</td>
</tr>
<tr>
<td>Research Design</td>
<td>The research design adopted is clear and understandable</td>
</tr>
<tr>
<td>Research Design</td>
<td>(Investigator bias—only for quantitative studies examining causation) The document describes the incorporation of a blinded path between investigator and data source or data analysis</td>
</tr>
<tr>
<td>Measurements</td>
<td>The document reports how tools and measures were developed or adopted and validated</td>
</tr>
<tr>
<td>Measurements</td>
<td>In qualitative studies, steps were taken to confirm the researcher’s interpretations, e.g., validation by study subjects</td>
</tr>
<tr>
<td>Measurements</td>
<td>Important variables/concepts used are defined precisely in the document, or sources are cited</td>
</tr>
<tr>
<td>Measurements</td>
<td>(Only in simulation-based studies) Analyses include basing input data on ecologically valid sources, validating findings in a real-life context, or both</td>
</tr>
<tr>
<td>Analysis</td>
<td>The practical significance of findings is clearly articulated in the document</td>
</tr>
<tr>
<td>Reporting</td>
<td>Appropriate sources were cited within the text and complete citations included at the end of the document</td>
</tr>
</tbody>
</table>

improved patient satisfaction, patient ratings of the location of the monitor, the ability to look at the screen, and perceptions regarding reviewing their medical records and other information are discussed. In the proposed surrogate estimation framework, the rating of this publication is level III**** (an independent study of high scientific strength and estimated high quality).

A second example is a qualitative study that explored what flexibility means to the caregivers and support personnel on an inpatient unit, and the domains of design decisions that potentially influence operational flexibility (Pati et al., 2008). Because of its qualitative research design, the study falls within the level VI class of evidence (Table 1). The article was published in a peer-reviewed journal (Environment & Behavior). The study was conducted within the flexibility framework, which is presented in the introduction. The three research questions are presented in a separate section. Sampling strategy (purposive) and sampling size are explained. The steps involved in data collection and analysis are clearly articulated. The development and testing of the interview tool are explained. Subjects’ validation of the researchers’ interpretations is reported. Precise definitions are provided for all important concepts. The practical implications of the findings are reported separately within each of the nine domains of design.
decision making found to influence operational flexibility. This study is rated as level VI***** (an independent study of moderate scientific strength and of estimated high quality).

**Discussion**

This paper addresses a contemporary challenge in EBD: the lack of a commonly agreed-on mechanism to evaluate evidence for use in design decision making. Among the key areas of challenge is the gap between the producers of scientific evidence and its intended consumers. It is hoped that the framework described in this article is simple enough to apply in practice without sacrificing the elements essential for the thorough scientific evaluation of evidence. It is anticipated that the framework will be critiqued and modified over time in a systematic, evolutionary process within healthcare and other building sectors.

**References**


GUEST EDITOR’S COLUMN
It’s All About Healing
Upali Nanda

RESEARCH METHODS COLUMNS
Research Methods for Design Practitioners To Evaluate Facilities
Mardelle McCuskey Shepley

Rooftop Hospital Gardens for Physical Therapy: A Post-Occupancy Evaluation
Brad E. Davis

EDITOR’S COLUMN
Ten Suggestions for Increasing Rigor in Architecture and Design
D. Kirk Hamilton

RESEARCH
Lactation Space Design: Supporting Evidence-Based Practice and the Baby-Friendly Hospital Initiative
Tammy Smith Thompson and LaShawna Heflin

A Comparison of Parent Satisfaction in an Open-Bay and Single-Family Room Neonatal Intensive Care Unit
Dennis C. Stevens, Carol C. Helseth, M. Akram Khan, David P. Munson, and E. J. Reid

Influence of Positive Distractions on Children in Two Clinic Waiting Areas
Debajyoti Pati and Upali Nanda

LETTERS TO THE EDITORS
Kathy Hathorn
Mary Grehan
Samira Pasha

BOOK REVIEW
Transforming the Healthcare Experience Through the Arts
Reviewed by Jill Sonke