Crossing the Evidence Chasm: Building Evidence Bridges from Process Changes to Clinical Outcomes

DAVID C. KENDRICK, MD, MPH, DAVIS BU, MD, MA, ERIC PAN, MD, MSc, BLACKFORD MIDDLETON, MD, MPH, MSc

Abstract

Objective: Although demand for information about the effectiveness and efficiency of health care information technology grows, large-scale resource-intensive randomized controlled trials of health care information technology remain impractical. New methods are needed to translate more commonly available clinical process measures into potential impact on clinical outcomes.

Design: The authors propose a method for building mathematical models based on published evidence that provides an evidence bridge between process changes and resulting clinical outcomes. This method combines tools from systematic review, influence diagramming, and health care simulations.

Measurements: The authors apply this method to create an evidence bridge between retinopathy screening rates and incidence of blindness in diabetic patients.

Results: The resulting model uses changes in eye examination rates and other evidence-based population parameters to generate clinical outcomes and costs in a Markov model.

Conclusion: This method may serve as an alternative to more expensive study designs and provide useful estimates of the impact of health care information technology on clinical outcomes through changes in clinical process measures.


The announcement\(^1\) and reaffirmation\(^2\) of the federal commitment to advancing health care information technology (HIT) has been further bolstered by events in the Gulf South after the recent hurricane seasons.\(^3\) This commitment creates both opportunities and challenges for health services and clinical informatics researchers. Clinicians, policy makers, lobbyists, economists, and the media demand evidence-based recommendations for HIT. To make decisions that will affect millions of lives and billions of dollars, decision makers require more than efficacy studies—they require results that indicate both the effectiveness and the efficiency of HIT solutions. The ability of the informatics research community to respond to this need with useful and credible evidence will determine our relevance to the debate.

Many evaluations of health services focus primarily on process measures.\(^4\) For example, there are numerous studies in the disease management literature that report the impact of technology on the rate of annual eye or foot examinations for diabetic patients.\(^5\)–\(^14\) However, there are few published studies that evaluate HIT's impact on the rate of blindness or amputations. Despite the increasing demand for credible clinical outcomes evidence, many studies in HIT lack the power to detect changes in clinical outcomes, a product of limited time and resources.\(^15\) In addition, the rapid evolution of new technologies makes the study subject itself, HIT, a moving target. By the time a large-scale trial is completed, the state of the art will have moved on.\(^16\)

Evaluations in HIT therefore tend to be relatively brief studies comparing convenient measures, more often made in a laboratory environment, or potentially idiosyncratic academic environments, than in real-world clinical settings, thus limiting generalizability. These studies would be classified by Fuchs and Garber\(^17\) as stage 1 and 2 technology assessments—evaluating the performance of the technology itself and perhaps the impact of the technology on processes of care. However, the demand for outcomes evidence mandates that future HIT research be at the level of stage 3 technology assessments, in which comprehensive clinical, economic, and social outcomes are evaluated to determine both the effectiveness and efficiency of the intervention. The importance of linking process measures to clinical outcomes has been previously described, but progress has been limited.\(^18\)
We propose an approach to maximize the ability of HIT evaluation research to report clinical and financial outcomes.

**Background**

The Center for Information Technology Leadership uses systematic review techniques and mathematical modeling to synthesize the best available evidence and quantify the value of HIT along clinical, financial, and organizational axes. Our most recent work assesses the value of information technology-enabled diabetes management (ITDM) programs. The general approach to this analysis has been to: (1) develop a population-based disease model, and (2) inform this model with results from studies of disease management interventions and programs. A comparison of clinical outcomes and disease costs before and after application of ITDM can thus be made.

The disease model is constructed as an influence diagram (ID) using Analytica software from Lumina Decision Systems, Inc. (Los Gatos, CA). This software enables the use of both qualitative and quantitative information and incorporates probability distributions to provide explicit tracking of uncertainty in the source evidence.

An issue arose in the course of this analysis that prompted us to develop a modeling method that we believe to be generalizable to other HIT evaluations. As is common in HIT research, much of the evidence found in the ITDM literature is reported as process changes, such as increases in the screening rates for diabetic retinopathy, peripheral neuropathy, and microalbuminuria, rather than impact on clinical outcomes, such as morbidity, mortality, and quality-adjusted life years. Clearly, evidence demonstrating impact on clinical outcomes is preferred. In fact, our disease management model relies on evidence demonstrating impact on disease and condition-specific clinical outcomes to drive its predictive engine. Thus, we found ourselves confronting a common question in HIT research: Can we translate evidence that demonstrates the impact of HIT on process changes into credible estimates of their impact on clinical outcomes?

Because the ideal evidence linking HIT and clinical outcomes is often not available, it is important to find alternative means of linking process changes to the clinical outcomes. We used evidence synthesis methods and modeling techniques to build logic pathways between process measures and clinical outcomes. We named these pathways evidence bridges, and by using them, we were able to derive estimates of clinical outcome from changes in process measures. Figure 1 provides a conceptual view of our solution to this problem. A method for constructing these bridges and an illustrative application are discussed below. In the example, we will construct an evidence bridge to link screening rates for diabetic eye disease with the incidence of blindness. Diabetic eye disease is the leading cause of new blindness among adults 20 to 74 years old, and each year 12,000 to 24,000 Americans with diabetes go blind.

**Model Formulation Process**

The general method for building evidence bridges is outlined here and is shown in Figure 2.

1. Identify the scope of the proposed evidence bridge, by defining the inputs (process measures that are available for modeling) and the outputs (clinical and financial outcomes of interest).
2. Propose a draft model, and list the key evidence required to populate the model.
3. Conduct a literature review to find each of the key pieces of evidence.
4. Assemble the model by adding the evidence and forming linkages.
5. Perform internal model refinement through stability testing and sensitivity analysis.
6. Assess external model validity by comparison with existing trials and datasets.
7. Formulate a model maintenance strategy.

We discuss each step in turn and provide an example from a model for diabetic eye disease.

**Scoping the Evidence Bridge**

The first step in creating the evidence bridge is to determine the goal and working parameters for the model. Essentially, this means identifying the inputs and outputs for the model that will become the bridge. It is often useful to start with a listing of the desired outputs. Although these may be fairly generic initially, such as “mortality” or “quality-adjusted life years,” it will be important to define these outcomes much more narrowly before beginning the modeling phase. It will
be tempting to define many outcomes of interest—this is an acceptable approach as long as the researcher understands that this list will need to narrow significantly as the process continues.

In our example, diabetic eye disease, we started with a range of outcomes of interest. However, it quickly became obvious that mortality due to diabetic eye disease is an exceedingly rare event and probably not an appropriate outcome to target. Estimating morbidity, therefore, became our objective. We further narrowed our focus to cases of blindness, the most widely studied outcome of diabetic eye disease.

Once the objectives of the evidence bridge have been defined, it is important to take stock of the data available for primary inputs to the model. In most HIT studies, these primary inputs will be changes in process measures impacted by the technology. Commonly, decreased chart pulls, increased number of user log-ins, and improved compliance with medications are reported. It is important to note that in order to be useful, these process changes need to have a documented baseline, and ideally, some estimation of the error associated with the measurements. Other important inputs include the baseline population and the prevalence of outcomes of interest in this population. In the case of most individual HIT studies, the initial model population should mirror the original study population in its characteristics. In situations in which this is not possible, sensitivity analysis and uncertainty estimation will be very important.

In the case of the diabetic eye disease evidence bridge, we needed to incorporate, and therefore normalize, process change inputs from several sources. The evidence was screened to determine the relative compatibility of the reported measures. For example, a variety of measures were reported for screening in diabetic eye disease, including compliance rate with annual eye examination, compliance with semiannual eye examination, and examinations per person-year of diabetes. The most commonly reported measure, compliance rate with annual eye examination, was chosen as the primary process measure input for the model.

**Drafting the Model**

With the major inputs and outputs of the evidence-bridging model defined, it is possible to begin the conceptual formulation of the model itself. Although there are several methodologies for creating models of this type, we favor the technique of influence diagramming. IDs were first described in the 1960s and have matured into a useful tool for decision analysis.

The creation of the draft model begins as in Figure 3, with the placement of variable nodes for the process changes (the inputs) at the left side of the diagram and objective nodes for the targeted clinical outcomes (the outputs) at the right side of the diagram.

Next, participants in the ID modeling process suggest concepts that they feel play a role in the hypothetical chain of logic linking the process measure and the clinical outcome.

**Figure 2.** Process diagram for constructing an evidence bridge.

**Figure 3.** The first step in scoping an evidence bridge is to determine the desired outcomes (outputs) and the available process measures (inputs).
For example, if the input is the screening rate for diabetic retinopathy, and the desired outcome is the incidence of blindness, one can immediately conclude that, in addition to the baseline rate of screening, the following data will be important: the sensitivity and specificity of the screening technique, the provider’s response to a positive test (percent of patients receiving phototherapy), the relative effectiveness of the treatment, and patient compliance with the treatment itself. In the diabetic eye disease model, the initial brainstorming sessions yielded a wide field of variables (shown in Figure 4).

Next, arrows indicating influence are placed on the diagram, linking the elements that are logically related to one another. These arrows are directional, and by convention, the arrow points from the influencer to the influencee (from conditioning event to conditioned event). For example, the number of persons screened, and the screening sensitivity and specificity will combine to influence four different nodes—true-positive, true-negative, false-positive, and false-negative. Similarly, true-positive and false-positive will combine with the treatment rate to influence the number of persons treated. Often, this process will elicit new insights into the relationships among the key parts of the system. It may become clear that an additional step is required to link two nodes. Equally likely is the possibility that some nodes, previously thought important, can be discarded. A somewhat subtle but important characteristic of an ID is that the absence of linkages between nodes, either direct or indirect, is an indication of conditional independence. This process of pruning and augmenting may occur over several cycles, and will ultimately yield the draft model.

In our example, linking variables with relationship arrows led to important new insights and the subsequent addition of several new variables to the draft model (Fig. 5). The draft model and initial list of variables helps to create an evidence table, which will guide the formal evidence review. A listing of the key variables for the diabetic eye disease model was assembled (Table 1), providing key words and concepts to guide the literature search to be undertaken in the next step.

**Identifying the Evidence**

Traditional sources of medical evidence, such as Medline and EMBASE, will naturally be used for the initial searches. Much of the best evidence will be found in these sources. However, for certain kinds of evidence it can be difficult to find any, much less ideal, sources. To limit the search to these traditional sources would be a mistake. Often, the information required to complete the evidence-bridging model will fall victim to publication bias—key evidence may not make it into publication because of negative results or an assumption that the results are too minor or inappropriate for the peer-reviewed medical literature. Consider our previous example, in which a link between the rate of screening for diabetic eye disease and the progression to blindness is sought. The sensitivity and specificity of the screening modality will play a key role, and is more likely to be found in a laboratory manual or a white paper than in the peer-reviewed medical literature. A model linking patient wait-times with costs due to lost working time might require information on wages and working hours from the Bureau of Labor Statistics. Other useful sources for population and health system statistics include the Census Bureau, Centers for Disease Control and Prevention, and National Center for Health Statistics.23 Other useful sources for population and health system statistics include the Census Bureau, Centers for Disease Control and Prevention, and National Center for Health Statistics.

**Figure 4.** Key variables in the diabetic eye disease system are assembled through brainstorming and discussion. The outcome of interest, blindness due to diabetic eye disease, is shown in a hexagon. DR = diabetic retinopathy.

**Figure 5.** Relationships have been added to the diagram showing the perceived direction of influence. The process of adding relationships has prompted the addition of more elements to the model. Those shown in trapezoids are items of key evidence, which will be used to calculate the other nodes. DR = diabetic retinopathy; QALY = quality-adjusted life years.
for Disease Control and Prevention, and Centers for Medicare and Medicaid Services.24–26

Often, HIT studies will be designed to report process changes that are relevant to one or more clinical care guidelines. These published guidelines, and the evidence cited by them, represent another important source of evidence. The Department of Health and Human Services provides the National Guideline Clearinghouse at www.guideline.gov, which is an excellent resource.27 Unfortunately, many of the guidelines are based on consensus panel reports, rather than a clear chain of published evidence linking the recommendations to clinical outcomes.

Another useful source of evidence is the so-called “grey literature,” which represents nontraditional publications across many disciplines. It is known that a high proportion of controlled trial results are only published as a part of conference or meeting proceedings, and never find their way into the traditional literature base.28 Fortunately, much of this information can be searched in the British Library’s “Inside” database, which catalogues the conference proceedings from more than 100,000 worldwide events annually.29 Patent databases, such as that of the U.S. Patent and Trademark Office, can provide a wealth of information about specific techniques and technologies, often with references leading back into the traditional literature.30

One of the most exciting recent developments in health care modeling is the increasing number of research datasets being made publicly accessible for qualified research. It is difficult to overstate the potential value of this development to modeling in health care. Working exclusively with the peer-reviewed literature limits the available evidence to the specific variables (and methods of representing them) chosen by the authors. Access to the actual datasets supporting those publications provides a whole new set of possibilities, as modelers can not only choose their own statistical methods to apply, but can formulate their own questions. For example, a published article may offer only an average and a standard deviation for an important variable. However, using the dataset behind the article, the modeler can go well beyond this single data point to create a complex polynomial that more closely represents the behavior of the variable, or limit analysis to an important subset of the population, or even evaluate the behavior of the variable with respect to other important variables in the dataset.

The initial evidence review likely will fail to uncover some key evidence. In this event, several options must be considered. In addition to a more aggressive literature search and review, an expert assessment can be applied. In this process, experts are queried for an estimate of the value in question, using a well-described methodology such as the modified Delphi method.31

It is likely that knowledge gained in the literature review will challenge key assumptions made in the creation of the draft model. New information could lead to an improved understanding of the system being modeled. For example, a new disease state, screening method, or population characteristic could be identified as an important contributor to the outcome of interest. In this case, alterations might be made to the draft model to accommodate these findings.

In the diabetic eye disease model, the search for each of the key pieces of evidence was initiated in PubMed. An article was found that described a financial model of diabetic eye disease made in the early 1990s by Javitt et al.32 The bibliography of this initial article led us to several additional articles that ultimately yielded much of the key evidence for our model.33–41 In addition, the literature review further informed our understanding of the disease process. In our initial draft model of diabetic eye disease, we considered only one disease state: diabetic retinopathy. However, the literature review revealed that macular edema caused significant morbidity, particularly among younger patients. Disease states for macular edema, and its logical extension, macular edema and diabetic retinopathy, were added to our model (Fig. 6). A representation of the completed retinopathy evidence bridge is shown in Figure 7.

We have found it useful to treat the draft model as such, and to actively seek alternative pathways to the conclusions of interest even if the model structure will need to be changed. Thus, a feedback loop has been included in the bridging model process diagram to indicate that adjustments and changes to the draft model may occur regularly throughout the evidence search and synthesis process.

Assembling the Model

Once the key evidence has been found and the model draft has been refined to include concepts prompted by the evidence search, the model assembly process can begin. The primary task in this step is to leverage mathematical techniques to incorporate the evidence in a way that allows the seamless connection of concepts to yield a functioning model. Units of measure must be normalized and the variety of uncertainty estimates converted to a standard for easy comparison. Also, in the event that multiple sources for evidence are found, the best evidence must be selected or several items of evidence combined where possible. Each modeling task brings with it unexpected challenges that may require creativity and novel approaches to the data.

In our diabetic eye disease evidence bridge example, the model assembly process, and in particular, the incorporation of the evidence into the draft model, presented several interesting challenges. The first required standardization of units to accept data from a variety of studies. The second related to the selection of an appropriate baseline for the

---

**Table 1** Initial Key Evidence Table

<table>
<thead>
<tr>
<th>Key Evidence</th>
<th>Measure Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence of background retinopathy</td>
<td>%</td>
</tr>
<tr>
<td>Diabetic retinopathy incidence</td>
<td>%</td>
</tr>
<tr>
<td>Screening cost</td>
<td>$/screening</td>
</tr>
<tr>
<td>Screening sensitivity</td>
<td>%</td>
</tr>
<tr>
<td>Screening specificity</td>
<td>%</td>
</tr>
<tr>
<td>Treatment rate</td>
<td>%</td>
</tr>
<tr>
<td>Treatment cost</td>
<td>$/procedure</td>
</tr>
<tr>
<td>Treatment success rate</td>
<td>%</td>
</tr>
<tr>
<td>Annual untreated blindness rate</td>
<td>%</td>
</tr>
<tr>
<td>Annual blindness cost</td>
<td>$</td>
</tr>
<tr>
<td>Quality-adjusted life years for blindness</td>
<td>Fraction of year</td>
</tr>
</tbody>
</table>

It is useful to construct a key evidence table to guide the literature review and evidence synthesis process. Shown here is the initial key evidence table for the diabetic eye disease bridging model.

---

The initial evidence review likely will fail to uncover some key evidence. In this event, several options must be considered. In addition to a more aggressive literature search and review, an expert assessment can be applied. In this process, experts are queried for an estimate of the value in question, using a well-described methodology such as the modified Delphi method.31

It is likely that knowledge gained in the literature review will challenge key assumptions made in the creation of the draft model. New information could lead to an improved understanding of the system being modeled. For example, a new disease state, screening method, or population characteristic could be identified as an important contributor to the outcome of interest. In this case, alterations might be made to the draft model to accommodate these findings.

In the diabetic eye disease model, the search for each of the key pieces of evidence was initiated in PubMed. An article was found that described a financial model of diabetic eye disease made in the early 1990s by Javitt et al.32 The bibliography of this initial article led us to several additional articles that ultimately yielded much of the key evidence for our model.33–41 In addition, the literature review further informed our understanding of the disease process. In our initial draft model of diabetic eye disease, we considered only one disease state: diabetic retinopathy. However, the literature review revealed that macular edema caused significant morbidity, particularly among younger patients. Disease states for macular edema, and its logical extension, macular edema and diabetic retinopathy, were added to our model (Fig. 6). A representation of the completed retinopathy evidence bridge is shown in Figure 7.

We have found it useful to treat the draft model as such, and to actively seek alternative pathways to the conclusions of interest even if the model structure will need to be changed. Thus, a feedback loop has been included in the bridging model process diagram to indicate that adjustments and changes to the draft model may occur regularly throughout the evidence search and synthesis process.

**Assembling the Model**

Once the key evidence has been found and the model draft has been refined to include concepts prompted by the evidence search, the model assembly process can begin. The primary task in this step is to leverage mathematical techniques to incorporate the evidence in a way that allows the seamless connection of concepts to yield a functioning model. Units of measure must be normalized and the variety of uncertainty estimates converted to a standard for easy comparison. Also, in the event that multiple sources for evidence are found, the best evidence must be selected or several items of evidence combined where possible. Each modeling task brings with it unexpected challenges that may require creativity and novel approaches to the data.

In our diabetic eye disease evidence bridge example, the model assembly process, and in particular, the incorporation of the evidence into the draft model, presented several interesting challenges. The first required standardization of units to accept data from a variety of studies. The second related to the selection of an appropriate baseline for the
process measures. The third involved the definition of a mathematical transformation for incorporating results from various studies. Finally, the last involved the selection of the best evidence where multiple sources existed. These challenges and the solutions described below should be generalizable to other modeling situations.

Careful attention should be given to the units of measure reported in the medical literature. For example, data reported in terms of events per person-year may need to be converted to annual events in order to be combined later in the model with a duration (in years) to yield the total number of events. Although the units often appear clear, in some cases, it may be difficult to determine which people are included in the denominator. For example, some studies may report a result as events per diabetic patient, whereas others may report events per study participant or even events per diabetic patient, with a different definition of diabetes. Detailed information about the definitions of the measures will be required to convert these results to other units. In some cases, correspondence with the author may be required.

It is important, where possible, to include measures of uncertainty for each variable in the model. Because evidence is derived from a variety of sources, rather than a single study, it is useful to have some sense of the significance of each data point, as well as the cumulative uncertainty accruing in the model. For the most uncertain estimates, this information can guide the researcher to seek better evidence, to revise the model to work with alternative evidence, or simply to provide a quantifiable note of caution when communicating the model results.

The goal of our diabetic eye disease model is to determine the relative impact of various technologies on the outcome of interest, cases of blindness. Our evidence-bridging model needed to accept, from many different studies, quantifiable measures of impact on the odds of receiving an annual dilated eye examination. Because each study reported im-

**Figure 6.** (a) The initial model conceptual diagram showing the disease states of interest. (b) The model was updated to include the additional disease states after the initial literature review revealed them to be important. DR = diabetic retinopathy, ME = macular edema, Post-Tx = posttreatment state.

**Figure 7.** Fully described retinopathy evidence bridge.
impact as a comparison with its own internal baseline (control group), the model required an annual screening rate that would reflect the baseline in the community at large. McGlynn et al. have shown a dramatic difference between the care recommended by clinical guidelines and the care actually delivered to patients in the United States. McGlynn estimates that only 14.21% of all U.S. diabetic patients receive their annual eye examinations. This estimate was adopted as the baseline annual eye examination rate in the model. The National Health and Nutrition Examination Survey also provides an excellent source of baseline information, as its data are carefully gathered to represent an accurate distribution of the U.S. population.

As noted above, studies of IT in disease management often report a percentage change in the rate of eye examinations. The magnitude of this change depends heavily on the baseline rate of eye examinations in the study population. For example, in a community where the baseline eye examination rate is 25%, it is believable that the rate could increase by 300% to a 75% rate of annual eye examinations. However, the same level of impact could not possibly be experienced in a community where the baseline eye examination rate is already 50%, because of the impossibility of screening more than 100% of the eligible population. Thus, we needed a mathematical transformation that allowed us to compare the impact from the two studies without passing impossible (i.e., >0% and >100%) screening rates into the model.

To address this issue, we changed our perspective on the metrics. Rather than consider the increase in screening, we took the opposite approach and focused instead on the decrease in the lack of screening. This enabled us to use a relative reduction in the lack of screening to alter the baseline screening rate in our model, and to be certain that the calculated screening rate would not exceed 100%. It is possible that a dramatically negative impact on screening, that is, an increase in the lack of screening, could push the screening rate yielded by this function below zero. However, the evidence did not support this possibility, and this method was incorporated into the model.

In some instances, the evidence base may yield several estimates of interest, and a decision will have to be made to select a specific piece of evidence or, alternatively, to combine the evidence into a single summary statistic. Systematic review and meta-analysis provide several tools for accomplishing this.

If the studies are deemed to be sufficiently similar (with regard to population, design, intervention, etc.) and report results that are numerically compatible, a meta-analytic statistical method, such as fixed or random-effects models, may be used to combine the results into a single effect size. When appropriately applied, this value will represent the overall magnitude of the effect and will provide some estimate of uncertainty as well. In general, the availability of more than five estimates from similar studies supports the use of these meta-analytic methods.

In the event that the multiple evidence points are derived from widely varying studies or are reported in numerically incompatible formats, a quality scoring method can be used to identify which study provides the “best” evidence. Although the specific application of a quality scoring method will vary in each situation, in general, the process should be logical and reproducible and should take into account the characteristics most important to the model’s accuracy. For example, in the case of the diabetic eye disease model, we reviewed dozens of studies to determine the impact of various interventions on the compliance with annual screening examinations. In many cases we found multiple point estimates, but generally fewer than four. Thus, we applied a quality scoring method. The objective of quality scoring was to rank order the studies themselves, not the point estimates under consideration. A score sheet was devised that emphasized the characteristics of the study most important to our modeling situation: study design, study population, study year, study duration, whether results were peer reviewed, presence of confounding factors, and fit to the research question. The score sheet and the original articles were provided to a minimum of three reviewers and each independently scored the studies. The point estimates from the highest-scoring studies were used in the model, and the range of the remaining estimates provided the boundaries for sensitivity analysis. A kappa statistic, a measure of the agreement between reviewers, often is appropriate to report as part of the final study results.

Refining the Model

Once the model is assembled and producing results, it is important to evaluate the model for internal validity, such as through stability testing and sensitivity analysis.

If the model is stochastic, it will be important to determine the model’s stability, or the precision of its results from run to run. In the case of Monte Carlo simulations, for example, it will be important to specify the level of precision required in the outcome results, which in turn enables the calculation of sample sizes required. The diabetic eye disease bridging model, a Monte Carlo simulation, required stability testing. In particular, this meant identifying the number of iterations required to overcome noise in the model and achieve results replicable within a predefined tolerance for run-to-run variance. We chose the standard error divided by the mean as the measure of variance. Focusing on the rate of blindness as the outcome of interest, we chose a stability threshold of 1%—that is, stability was said to be achieved when the run-to-run variance of the rate of blindness dropped below 1%. This performance was achieved with 40,000 iterations.

As with many modeling decisions, the selection of 1% variance as the threshold for acceptable stability was made based on the frequency of the events of interest. In this situation, we thought that the model should be capable of detecting a 1% difference between the impacts of various screening interventions on the rate of blindness.

Next, sensitivity analyses are performed to identify the variables most influential in the model’s outcomes. This information can be used to further refine the model. For example, if the sensitivity analysis reveals that a variable from a relatively low-quality or suspect source is driving the results, improving the credibility and accuracy of the estimate becomes important. This may involve further literature review or even conduction of a mini-experiment to arrive at a better estimate.
Sensitivity analyses show that variation in the overall incidence of diabetic eye disease has the widest impact on blindness (Fig. 8). Should a better estimate for the incidence of diabetic eye disease become available, it can be included in the model. The sensitivity and specificity of the screening test are both influential, highlighting the importance of continued improvements in screening methods. Worsening posttreatment outcomes result in significant increases in blindness, but similar improvements do not result in a commensurate decrease in blindness.

Validating the Model
The gold standard for evaluating simulations is, of course, to compare their performance to reality. The best proxy for reality that we are aware of is the appropriately conducted randomized controlled trial (RCT), in which the populations are well characterized, the patients’ environments are most carefully standardized, and the interventions are applied uniformly to the test group. When such trials are available for comparison, a sufficiently robust model could be tested to determine its ability to simulate the trial(s). If possible, a model should be validated against multiple trials, because in general, confidence in a model’s predictions grows as it is validated against more and more diverse trials.

In most cases, it seems unlikely that appropriate RCTs will be found for the validation of evidence-bridging models. Alternatives to validation by RCT simulation include the following.

Unit Validation
Analogous to unit testing in software engineering, unit validation of a model involves validating subsections of the model that correspond to outcomes of known trials or datasets. For example, a function representing treatment compliance may have been derived from an RCT. To the extent possible, the parameters of the RCT should be set in the model and the value of treatment compliance should be monitored to ensure that it generates results similar to the RCT.

Boundary Validation
A common approach to solving problems in mathematics involves testing the upper and lower bounds of a function. This approach also can be applied to validate a model. In the case of the diabetic eye disease bridging model, we are able to validate the baseline case, in which the model is configured for present conditions (screening rate 14.21%, sensitivity 80%, specificity 97%). As shown in Figure 9, the rates of blindness tend downward with increased screening rates. In addition, the rate of blindness predicted for today’s state (baseline) is a 1.8% cumulative ten-year incidence, which is within 0.1% of the ten-year cumulative incidence rate proposed by the Centers for Disease Control and Prevention (a relative increase of approximately 6.4%). If other data were available, such as regional incidence of blindness and region-specific screening rates, the model could be further validated against these points, as long as the population could be tailored to the region as well.

Model Comparison
In some cases, models are compared against one another or against a common dataset to evaluate their relative accuracies. In particular this has been seen with the Mt. Hood challenge, in which diabetes models were cross-tested.

Data Splitting
Models built from datasets, rather than assumptions derived from published literature, can be validated using a technique called dataset splitting. During the construction of the model, the dataset is split into two (or more) parts using random sampling methods. The data in one part are used to build the model, whereas the data in remaining part(s) can be used in validation of the model.

The importance of model validation cannot be overstated. Indeed, the level of confidence placed in the results a model produces should be proportional to the degree to which the model has been validated. Admittedly, in the case of evidence-bridging models, there are likely to be few opportunities to validate against RCTs. However, the methods and
science of model validation are continuing to evolve, and it is hoped that the increasing availability of study datasets and improvements in methods will enable higher-quality validations.

Model Maintenance
Once constructed, an evidence-bridging model may be applied to many situations. For example, the diabetic eye disease bridging model may be used to compare many different interventions designed to encourage increased screening rates. Just as knowledge in health care is continually evolving, so too will the evidence base related to the bridging model. Depending on the importance of the model, it may become necessary to establish a process for updating, versioning, and validating the model to ensure that it remains consistent with medical science. Methods for accomplishing this are beyond the scope of this article, but can range from an informal monitoring of the medical literature to implementation of a complex editorial review and model revision and validation process. Ideally, the level of resources dedicated to the model maintenance effort should be commensurate with the value of the answers the model produces. For example, an evidence-bridging model linking the rate of eye examinations to clinical outcomes could be very important to a health plan designing a pay-for-performance plan, and should be well maintained.

Discussion
The difficulty of evaluating health care IT through randomized controlled trials that are sufficiently powered to report outcomes has been noted. The work of Mant and Hicks provides further insight into the plight of HIT researchers and the general need for an evidence bridge to link process measures to clinical outcomes. They have shown, through evaluations of disease-specific mortality from myocardial infarction (MI) as a quality metric for comparing hospitals, that the use of outcome measures to compare quality and performance may not be the most effective approach. Even when assuming an identical case identification and mix, they determined that it could take 73 years to detect a statistically significant 3% relative reduction in MI-specific mortality between two hospitals. In contrast, the use of accepted process measures, such as thrombolysis and the use of aspirin, beta-blockers, and angiotensin-converting enzyme inhibitors, yielded this determination with only four months of data.

An analogous situation must exist in health care IT—the pursuit of outcomes-level results in every HIT study may well be a fool’s errand. It seems logical, based on the example above, that the changes in process measures reported by HIT studies should be evidence enough of the differences between technologies and controls. However, in the case of MI, the process measures have been well studied and their impact on the outcome of interest has been quantified. In contrast, many of the process measures reported by HIT evaluations have not been quantitatively linked to the outcomes of interest. Evidence bridges provide a means of quantifying this linkage.

We have described an evaluation scenario particularly suited to the use of the evidence bridge method. Studies of various technologies report the effect on many different process measures, and the task of the evidence bridge here is to: (1) accept one or more common process changes, and (2) translate those process changes into the outcomes of interest. Evidence bridges could be created for many process measures that are readily measured and commonly reported in HIT research. These models could be gathered into a library to be used to link the results of any study reporting process outcomes to important clinical and economic outcomes. For example, once constructed and validated, the retinopathy bridging model can be used to predict the clinical outcomes of any intervention (technology or otherwise) that can be shown to improve annual eye examination rates. In the course of completing the Center for Information Technology Leadership diabetes model, similar evidence bridges were built for diabetic peripheral neuropathy and nephropathy, leveraging increases in foot examination rates and microalbuminuria screening to show an impact on clinical outcomes.

There are three additional scenarios that we think are particularly well-suited to the use of bridging models. In the first scenario, the evidence bridge could be useful in vetting
clinical guidelines. Despite efforts to support all clinical guidelines with solid evidence, a significant number of guidelines still contain recommendations based on expert opinion and consensus reports. Rather than conduct RCTs for each of these guidelines, it may be easier and more cost effective to construct models informed by the best available evidence to link the recommended clinical processes with the outcomes of interest. This method allows testing of a variety of scenarios and potentially the identification of the most appropriate levels of intervention. At a minimum, the evidence bridge will allow an objective assessment of the impact of variations in expert opinion, and the sensitivity analysis process could provide guidance for further research.

In a second scenario, the evidence bridge method can be used prospectively in the planning process for a research study. If the study is expected to yield process changes, and it is hoped that an evidence bridge may be built to show impact on outcomes, there are several reasons to build the evidence bridge before finalizing the study design. First, the process of building the bridge may yield new insights into the specific process measures that will be required. Second, the completed evidence bridge may indicate the process-change effect size required to make a meaningful change in the outcome. This will be important in sizing the study appropriately.

Finally, in a third scenario, the evidence bridge method can become a tool for implementing a national research agenda for HIT. Funding agencies could identify the outcomes most important for policy decisions and commission projects to construct standardized evidence bridges to provide these outcomes. Provided with an approved and validated set of evidence bridges, HIT researchers would be free to pursue shorter, lower-cost studies that focus on identifying impacts on process measures. More technologies, and their impacts in more populations, could be evaluated in a shorter time frame, at a cost of fewer resources. Because the evidence bridge in this scenario will potentially have much larger and more widespread use, it may be advisable to invest in primary research to inform the evidence bridge where the existing literature base is not convincing or robust. The process of building the bridge should be useful in identifying the variables most in need of a focused primary research study to inform them.

As with any modeling technique, this method has important limitations. The Center for Information Technology Leadership method makes use of the best available evidence to arrive at a complete model. Although the retinopathy bridging model was based entirely on evidence published in the peer-reviewed medical literature, in other instances the best evidence source may be a white paper or expert opinion. A model may inherit the limitations of the literature supporting it, and further, its generalizability may be limited because the combined evidence base derives from studies done across many, potentially different, populations.

Despite the limitations, the very process of modeling offers value. For example, working through the modeling process often brings new insights into key elements of a system, as with the inclusion of new disease states in the diabetic eye disease bridge model. At a minimum, missing evidence provides a clear pathway for future study, and at a maximum, it may force the modeler to reconsider initial perceptions and assumptions about the system in question.

**Conclusions**

The evidence bridge modeling method can be used to build evidence bridges between process changes and important outcomes, providing essential insight into the effectiveness and efficiency of HIT. By providing a common insertion point for process measure results from many different studies, these bridging models can be used to compare various HIT interventions on a level playing field. These models may not be perfect in the first iteration, but targeted investigations can be used to further clarify assumptions and improve accuracy.

**References**

15. Olivarius NF, Beck-Nielsen H, Andreassen AH, Horder M, Pedersen PA. Randomised controlled trial of structured per-