CHAPTER

12

BENCHMARKING CLINICIANS

Farrokh Alemi

In hiring, promoting, and managing clinicians, managers often need to understand the efficiency and effectiveness of clinical practices. To accomplish this, they need to benchmark clinicians, usually on the basis of risk-adjusted expected cost. This chapter describes how to organize benchmarking efforts.

Why Should Clinicians Be Benchmarked?

Managers are focused on the survival of their organization, which often translates to two overarching objectives: improve productivity and increase market share. Clinicians' decision making affects both the organization's productivity and its market share. Poor clinicians are bad for the patient and for the organization. Inefficient clinicians increase the cost of care for everyone. Clinicians with poor quality of care affect the reputation of the organization and, de facto, its market share. Managers who ignore poor quality of care among their clinicians and focus on nonclinical issues are failing to see the real causes of their organization's malaise. If a manager is serious about improving the long-term financial performance of her organization, she has no choice but to address clinicians' practice patterns.

For a long time, managers have avoided addressing the quality of clinical decisions on the grounds that they do not have sufficient training to understand these decisions and because such managerial interventions would be an unwelcome intrusion in the patient-provider relationship. But are these criticisms valid? Do managers need to know medicine to understand practice patterns?

Managers can profile physicians by looking at the outcomes of their patients. They may not understand how a patient should be managed but they certainly can understand patient outcomes such as mortality, morbidity, satisfaction, health status, and numerous other measures. Managers can then compare clinicians to each other and see who is performing better. Across encounters and over time, the manager detects patterns and uses

This book has a companion web site that features narrated presentations, animated examples, PowerPoint slides, online tools, web links, additional readings, and examples of students' work. To access this chapter's learning tools, go to ache.org/DecisionAnalysis and select Chapter 12.

this information to bring about lasting changes in practice patterns. Typically, the information is provided back to a group of clinicians who identify and propagate the best practices.

The concern that benchmarking intervenes in physician and patient relationships might be a red herring. After all, practice profiles are constructed after the fact, when the patient is gone. Practice profiles do not indicate how an individual patient should be managed; rather, they identify patterns across individual visits. In short, these profiles leave the management of individual patients in the hands of the physician. There is no interference in these clinical decisions. No one tells the clinician to prescribe certain drugs or to avoid certain surgeries for a specific patient. Practice profiles document the net effect of the physician on groups of patients; these profiles provide information about a clinician's performance overall.

Practice profiles can help patients select the best clinicians. Managers need to act responsibly about who they hire and promote and practice profiles can inform them. Providers can use practice profiles to learn from each other. Patients, managers, and providers can use practice profiles, if accurate profiles can be constructed and easily communicated to them.

How Should Benchmarking Be Done?

In benchmarking, a clinician's performance is compared to the expected outcomes of her peers. This expectation is set in many different ways. This section reviews some typical methods.

Benchmarking Without Risk Adjustment: Comparing Clinicians to the Average Performance of Peers

The most common benchmarking method is to compare a clinician to the average performance of his peers. A statistical procedure for the analysis of the means of two samples (mean of outcomes for the clinician and mean of outcomes for the peer providers) is well established. Excel software contains a tool for such analysis. Analysts can use these procedures to see if the difference in means is statistically significant.

An example may demonstrate this type of benchmarking. Callahan, Fein, and Battleman (2002) set out to benchmark the performance of 123 internal medicine residents at the New York-Presbyterian Hospital in New York City. The outcomes examined included the following:

- Patients' satisfaction as measured by telephone interviews of at least ten patients of the resident;
- Disease-management profiles for an average of 7 patients with diabetes and 11 patients with hypertension. Data reported included measures of patient's condition and frequency of use of various medications; and
- Faculty evaluations on seven dimensions, including
  - History taking
  - Physical examination
  - Differential diagnosis
  - Diagnostic and treatment plan
  - Healthcare maintenance
  - Compassion
  - Being a team player

Each resident received data about her individual performance compared with the mean data for his peer group (the remaining residents). Feedback was provided once every six months. Figure 12.1 shows an example of a report received by the residents. This report shows the mean for clinician X and the overall average of all remaining clinicians.

Callahan, Fein, and Battleman's (2002) benchmarking of residents was successful. The analysis identified a variety of performance patterns among residents. Some were above average and others below. Residents reacted positively to the feedback, and some considered it as the "most comprehensive evaluation" they had received to date.

But on the negative side, this benchmarking as well as all other efforts of comparing the unadjusted mean outcomes of clinicians, is flawed. It does not account for the severity of the illness of the patients of the clinicians. On the surface, all patients may have the same illness and can be categorized together; but in reality, some patients, even those with the same disease, are sicker than others. Logically, worse outcomes are expected for sicker patients. Clinicians treating sicker patients will have below average outcomes, not because of their performance but because they started with patients who had poorer prognoses. Naturally, many clinicians question benchmarking efforts on the grounds that these efforts do not adequately account for the differences in their patient populations. To avoid these errors and to address the concerns of clinicians regarding apple-to-apple comparisons, it is important to adjust for differences of case mix among the clinicians.

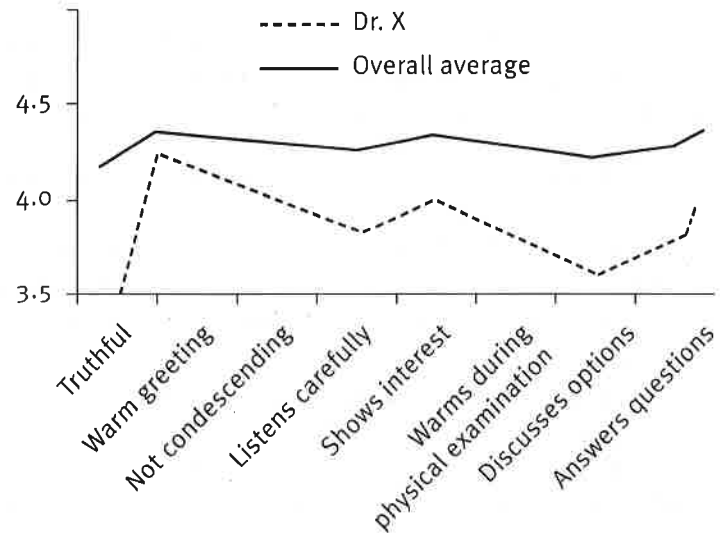

FIGURE 12.1 Comparing One Resident's Performance on Satisfaction with the Average of Other Residents

SOURCE: Callahan, M., O. Fein, and D. Battleman. 2002. "A Practice-Profiling System for Residents." *Academic Medicine* 77 (1): 34–9. Used with permission.

Risk-Adjusted Benchmarking: Comparing Clinicians to Peers on Patients of Similar Severity of Illness

Three methods of comparing clinicians on similar patients are presented here; the choice of which to use will depend on what data are available. In the first approach, the performance of the peer providers is simulated on the patients of the clinician being evaluated. First, the severity of each patient is noted and the probability of having patients in a particular severity group is calculated. If P_i is the probability of observing the patient in the severity group i, and $O_{i,\text{ clinician}}$ is the average outcome for the clinician for severity group i, then the following formula is used to calculate expected outcomes for the clinician, $O_{\text{clinician}}$:

$$O_{\text{clinician}} = \sum P_i \times O_{i,\text{ clinician}},$$

where i = low, medium, and high severity.

Next, the analyst projects the performance of peer providers on the type of patients seen by the clinician being evaluated. This is done by weighting the outcomes of peer providers according to the frequency of patients seen by the clinician being evaluated:

$$O_{\text{peer providers}} = \sum P_i \times O_{i,\text{ peer providers}},$$

where i = low, medium, and high severity.

Note that in the above formula, the probability of observing a patient in a severity category comes from the clinician's practice and not from the peer provider's practice.

For example, consider that a clinician and her peers have had the outcomes displayed in Table 12.1. Is this clinician better or worse than the peer providers? To answer this question, the analyst must compare the expected outcomes for the clinician to the expected outcomes for the peer providers simulated on the same patients as the clinician.

The first step is to calculate the probability of finding a patient in a different severity grouping. This is done by dividing the number of patients in a severity group by the total number of patients seen by the clinician being evaluated. The probability of having a low severity patient is 20/120, a medium severity patient is 30/120, and a high severity patient is 70/120. This clinician mostly sees severely ill patients. Once the probabilities are calculated, the second step is to calculate the expected length of stay of the patients of the clinician being evaluated:

$$O_{\text{clinician}} = (20 \div 120) \times 3.1 + (30 \div 120) \times 3.4 + (70 \div 120) \times 5.2 = 4.4 \text{ days.}$$

To understand if 4.4 days is too high or too low, the analyst needs to compare this clinician's performance to her peer providers. But the peer providers do not see patients who are as severely ill as those of the clinician being evaluated. To simulate the performance of the peer providers on the patients seen by the clinician, the analyst uses the frequency of severity among that clinician's patients to weight the outcomes of the peer providers:

$$O_{\text{peer providers}} = (20 \div 130) \times 4.1 + (30 \div 120) \times 3.0 + (70 \div 120) \times 4.5 = 4.1 \text{ days.}$$

The clinician whose data are being analyzed seems to be less efficient than the average of her peer group. Note that in both analyses the same frequency of having low, medium, and high severity patients is used. Therefore, the differences cannot be caused by the severity of patients. Of

TABLE 12.1 Severity-Adjusted Comparison of the Performance of Several Clinicians

	Clinician		*Peer Providers*	
Severity of Patients	*Number of Patients*	*Average Length of Stay of Patients*	*Number of Patients*	*Average Length of Stay of Patients*
Low	20	3.1	80	4.1
Medium	30	3.4	10	3.0
High	70	5.2	10	4.5

course, the analysis can be misleading if the classification of patients into various severity groups is faulty or if observed differences are caused by random variations and not by real practice differences. But if the classification of patients into severity groups is correct, the fact that the projected length of stay for peer providers was lower than the clinician suggests that peer providers may be more efficient.

Risk-Adjusted Benchmarking: Comparing Clinicians to Expected Prognosis

Another way to construct meaningful risk-adjusted benchmarks for a practice is to compare observed outcomes against what would be expected from patients' prognoses. A patient's prognosis can be estimated from the patient's severity on admission (as shown in Chapter 2 by the construction of a severity index), from the patient's self-reported expectation, or from the judgment of other clinicians. Once the patient's prognosis is known, the observed outcomes can be compared and variations from the expected prognosis can be noted. For example, assume that using the Acute Physiological Chronic Health Evaluation severity index you have predicted the expected length of stay of 30 patients to be as indicated in Table 12.2.

The first step in comparing the observed and expected values is to calculate the difference between the two. Then, the standard deviation of the difference is used to calculate the *Student's t-test.* The Student t-test is used to decide if the differences between observed and expected values are statistically significant. Excel provides a program for calculating the Student's t-test in paired observations. Assume that you have obtained the results shown in Table 12.3 by using this program.

The analysis showed that the length of stay of patients of this clinician were lower than the expected values on admission. Therefore, this clinician's practice is more efficient than the expectation.

Risk-Adjusted Benchmarking: Comparing Clinicians When Patient's Severity of Illness Is Not Known

As the previous two sections have shown, in any benchmarking effort it is important to make sure that you compare clinicians on the same type of patients; otherwise, it would be like comparing apples to oranges. Every provider sees a different patient. The typical approach is to measure the severity of the illness of the patient and build that into the analysis. For example, in the first approach, patients were divided into broad categories of severity (low, medium, and high) and care provided within each category was compared. In the second approach, patients' severity of illness was used to forecast their prognoses and compare this forecast to observed outcomes. Both methods are built on access to a reliable and valid measure of

TABLE 12.2 Expected and Observed Length of Stay

Case Number	Length of Stay		Difference	Case Number	Length of Stay		Difference
	Expected	Observed			Expected	Observed	
1	5	4	1	16	4	4	0
2	7	3	4	17	6	3	3
3	6	4	2	18	3	5	−2
4	4	3	1	19	7	5	2
5	6	4	2	20	3	5	−2
6	8	6	2	21	6	5	1
7	6	3	3	22	3	4	−1
8	4	4	0	23	5	4	1
9	4	3	1	24	3	4	−1
10	3	3	0	25	5	4	1
11	7	3	4	26	3	4	−1
12	4	4	0	27	5	4	1
13	5	3	2	28	3	3	0
14	6	3	3	29	4	4	0
15	4	4	0	30	5	5	0

TABLE 12.3 Comparison of Expect and Observed Values Using Student-Statistics in Excel

	Expected	Observed
Mean	4.80	3.90
Variance	2.10	0.64
Observations	30.00	30.00
Pearson Correlation	0.10	
Hypothesized Mean Difference	0.00	
Degrees of freedom	29.00	
Student's t-test	3.11	
$P(T \leq t)$ one-tail	0.00	
Critical one-tail	1.70	
$P(T \leq t)$ two-tail	0.00	
Critical two-tail	2.05	

the severity of illness. However, sometimes such measures are not available or available measures do not adequately measure the full spectrum of the severity of the patient's illness. This section provides an alternative method of benchmarking that does not require the availability of a valid and accurate severity index.

When no severity index is available, an analyst must still make sure that apples are compared to apples by matching the patients seen by different providers feature by feature. The expected outcome for the clinician being evaluated is calculated as

$$O_{\text{clinician}} = \sum P_{j,\ldots,m} O_{j,\ldots,m,\text{ clinician}} \quad \text{for all values of } j,\ldots,m,$$

where

- $j,\ldots,m$ indicates a combination of features j through m;
- $P_{j,\ldots,m}$ indicates the probability of these features occurring; and
- $O_{j,\ldots,m,\text{ clinician}}$ indicates the clinician's outcomes when these features are present.

The expected outcome for the peer providers is calculated in a similar fashion, with one difference:

$$O_{\text{peer providers}} = \sum P_{j,\ldots,m}\ O_{j,\ldots,m,\text{ peer providers}} \text{ for all values of } j,\ldots,m.$$

In this calculation, the probabilities are based on the frequency of features among the patients seen by the clinician being evaluated, but the outcomes are based on the experience of peer providers. By using this formula, the analyst is simulating what the expected outcomes would have been for peer providers if they had the same patients as the clinician being evaluated.

An example can demonstrate the use of this procedure. Table 12.4 shows 20 patients of one clinician and 24 patients of her peer providers. These patients were admitted to a hospital for myocardial infarction (MI). In each case, two features were recorded: existence of a previous MI and presence of congestive heart failure (CHF). Obviously, a patient with a previous MI and with CHF has a worse prognosis than a patient without these features. The analyst needs to separate outcomes for patients with and without specific characteristics.

An *event tree* can be used to organize the data. An event tree is a decision tree without a decision node. Each feature can be used to create a new branch in the event tree. For example, the event tree for the patients seen by the clinician is provided in Figure 12.2.

Using the data in Table 12.4, the analyst can group patients and calculate the probabilities and average cost for clinician's patients. Figure 12.3 shows the result.

The expected length of stay for the patients of the clinician being evaluated is 5.4 days. This is obtained by folding back the tree to the root node. Starting from the right (the highlighted area in the formula), each node is replaced with the expected length of stay:

$$\text{Expected length of stay} = [6 \times 0.65 + 5 \times (1 - 0.65)] \times (0.85) + (4 \times 1.0 + 0) \times (1 - 0.85) = 5.4.$$

Expected value for top node to the right

Expected value for bottom node to the right

TABLE 12.4 Patients of the Clinician and Peer Providers May Differ in Severity of Illness

Clinician's patients				*Peer Provider's Patients*			
Case No.	*Previous MI*	*CHF*	*Length of Stay*	*Case No.*	*Previous MI*	*CHF*	*Length of Stay*
1	Yes	Yes	6	1	Yes	Yes	6
2	Yes	No	5	2	Yes	Yes	6
3	Yes	Yes	6	3	No	Yes	4
4	Yes	Yes	6	4	No	No	3
5	Yes	Yes	6	5	No	Yes	4
6	Yes	No	5	6	No	Yes	4
7	Yes	Yes	6	7	Yes	Yes	6
8	Yes	No	5	8	Yes	Yes	6
9	Yes	Yes	6	9	Yes	Yes	6
10	Yes	No	5	10	Yes	Yes	6
11	Yes	Yes	6	11	Yes	Yes	6
12	No	Yes	4	12	No	No	3
13	No	Yes	4	13	No	Yes	4
14	No	Yes	4	14	No	Yes	4
15	Yes	Yes	6	15	No	Yes	4
16	Yes	Yes	6	16	No	Yes	4
17	Yes	Yes	6	17	No	Yes	4
18	Yes	No	5	18	No	No	3
19	Yes	No	5	19	Yes	No	5
20	Yes	Yes	6	20	Yes	Yes	6
				21	Yes	Yes	6
				22	Yes	Yes	6
				23	Yes	No	5
				24	No	Yes	4

Procedures for folding back a tree were described the Chapter 5. To simulate how the same patients would have been cared for under the care of peer providers, the event tree is kept as before, but now the average length of stay of each patient grouping is replaced with the average length of stay of patients of the peer providers. Table 12.5 provides the average length of stay of patients seen by peer providers.

If one combines the event tree from the clinician's patients and the outcomes of peer providers, the result is the graph in Figure 12.4.

The expected stay (obtained by folding back the tree in Figure 12.4) of the patients of the peer providers is 5.4 days. Thus, the clinician being evaluated and the peer provider have similar practice patterns. Note that if the event tree had not been changed to reflect the probabilities of patients seen by the clinician, the expected length of stay for patients seen by the provider would have been 5.75 days; this would have led to the erroneous conclusion that the clinician is more efficient than his peer. This example

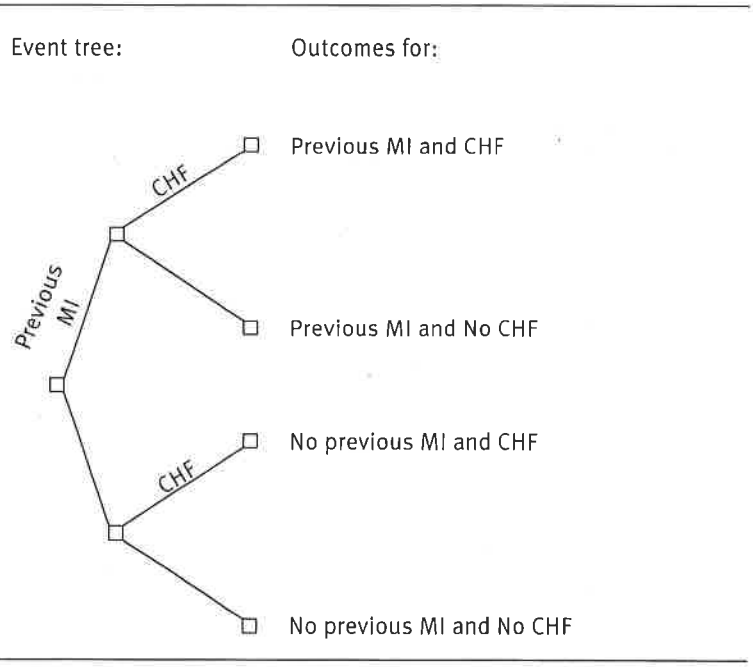

FIGURE 12.2 An Event Tree for the Clinician's Patients

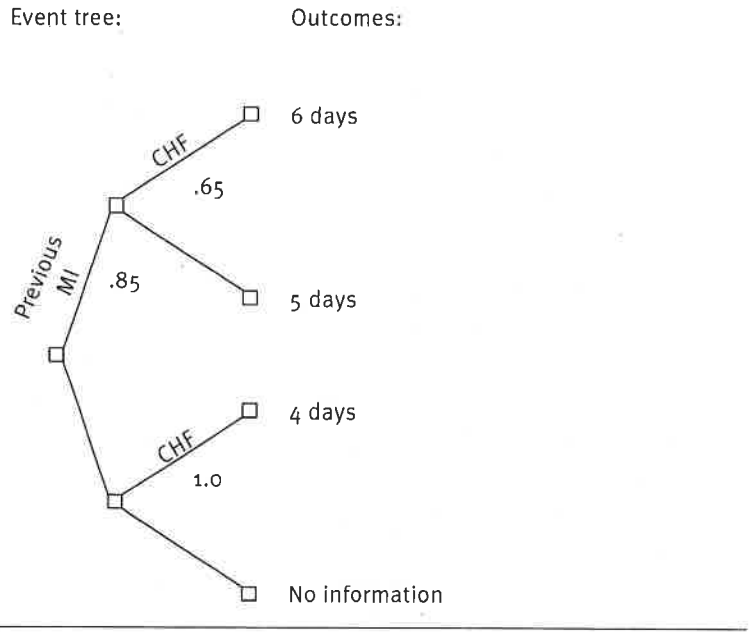

FIGURE 12.3 The Probability and Length of Stay for Different Patient Groups Seen by Clinician

highlights the importance of simulating the performance of peer providers on the patients seen by the clinician.

As the number of features increase, the number of data points that fall within each path on the decision tree becomes smaller. Soon, most

		Previous MI	
		No	*Yes*
CHF	*No*	3 days .13	5 days .08
	Yes	4 days .38	6 days .42

TABLE 12.5 Length of Stay and Probability of Observing Various Types of Patients Among Peer Providers' Patients

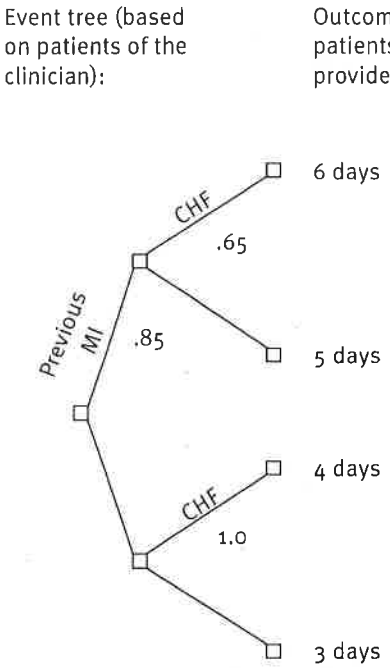

FIGURE 12.4 Projected Performance of Peer Providers on the Clinician's Patient

paths will have no patients. Many peer providers' patients cannot be matched on all features to the clinician's patients. When the features available do not match, the analyst can use expected outcomes to replace missing information. For example, consider if the only feature available among the clinician's patients was the presence of previous MI. No information was available on CHF for these patients. Now the event node available on the clinician's patient is as shown on the left of Figure 12.5. In addition, the outcomes for the peer providers need to be calculated as the expected outcomes. Note that from Table 12.5 and for patients of the peer providers,

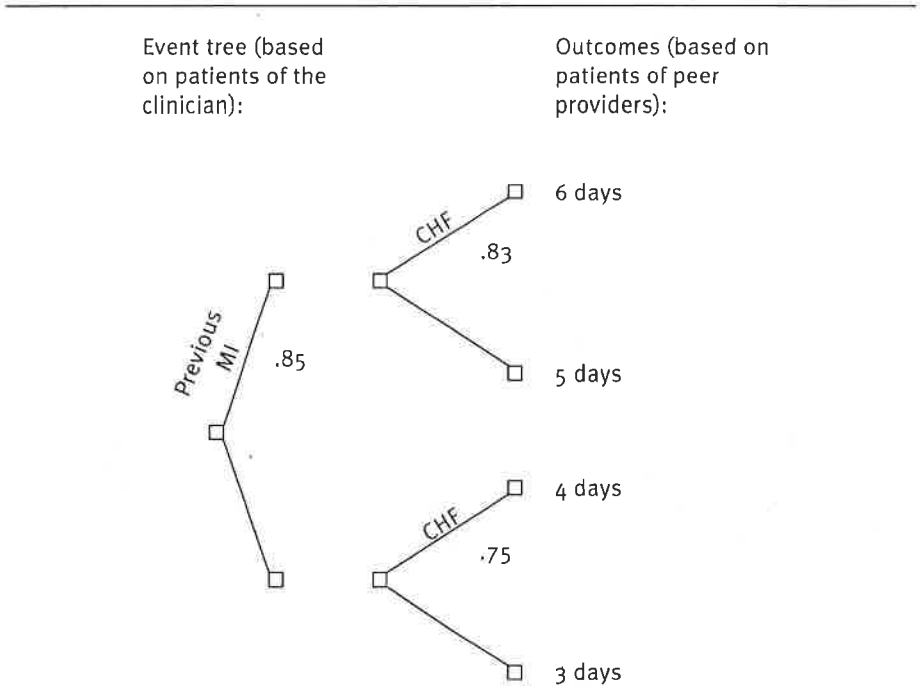

FIGURE 12.5 Projecting Peer Providers' Performance on the Clinician's Patient When Only "Previous MI" Feature Is Available

the probability of CHF for patients who have had a previous MI is calculated as

$$\text{Probability of CHF given a previous MI} = \frac{.42}{.08 + .42} = .83.$$

Similarly, the probability of CHF for patients who have not had a previous MI can calculated as

$$\text{Probability of CHF given no previous MI} = \frac{.38}{.13 + .38} = .75.$$

Figure 13.5 shows these probabilities and the outcomes used to calculate the expected outcome of the peer providers on patients seen by the clinician being evaluated.

The expected outcome for the peer providers is given as

$$O_{\text{peer providers}} = .85 \times .83 \times 6 + .85 \times (1 - .83) \times 5 + (1 - .85) \times .75 \times 4 + (1 - .85) \times (1 - .75) \times 3 = 5.5 \text{ days}.$$

One way to think about this procedure is that you have simulated the performance of peer providers on the patients of the clinician being

evaluated by replacing the probabilities with the corresponding values from the clinician's experience whenever such probabilities are available. When such values are not available, you should continue to use the probabilities from the experience of peer providers.

It is also possible that some features are found in the patients seen by the clinician being evaluated but not among the patients seen by the peer providers. In these circumstances, the expected outcome for peer providers is calculated on the basis of the event tree of the clinician being evaluated, and truncated to features shared among the clinician and the peer providers. Thus, if it is not clear among the peer providers' patients if they had CHF, then the event tree used to simulate the performance of peer providers is based on the experience of the clinician but without the CHF feature. In this fashion, the clinician and the peer providers can be compared even when features of some of the patients do not match the others' features.

When no features are shared between the patients seen by the clinician being evaluated and the patients seen by peer providers, the procedure provides an unadjusted comparison of the mean of the two groups. As the number of features shared between the two groups increases, there are more adjustments for the severity of illness of the patients.

What Are the Limitations of Benchmarking?

When the number of features on which patients of the clinician and peer providers must be matched increases, a visual display of the data through a decision tree is no longer feasible. Furthermore, because no two cases are likely to match on all features, strict feature-by-feature comparisons are not possible. In these situations, a modified approach is needed. One such modification is to weight patients in the clinician's care according to the similarity of these patients to the peer provider cases. Tversky (1977) proposed a method of assessing the similarity of two patients by examining features they share and features they do not share. Alemi, Haack, and Nemes (2001) used Tversky's approach to compare patients with similar, but not exactly the same, features.

When, because of a large number of features, decision trees are not used for benchmarking clinicians, they remain useful as a method of explaining how the analysis was done. The tree structure helps clinicians get a sense that like patients are being compared to each other. It reassures them that the analysis has followed their intuitions about comparing similar patients to each other.

Is it Reasonable to Benchmark Clinicians?

Risk assessment is not as benign as it first looks. When a clinician's performance is measured and the clinician is provided with feedback, several unintended consequences may occur:

1. *Measurement may distort goals.* Clinicians may improve their performance on one dimension but inadvertently deteriorate on another. For example, if the manager emphasizes length of stay, clinicians may improve on this measure but inadvertently increase the probability of rehospitalization because they have sent patients home too early. People tend to focus on what is measured and may ignore other issues. To avoid this shortcoming, it is important to select the benchmarking goals broadly and to select multiple benchmarks.
2. *Measurement may lead to defensive behavior.* Clinicians may put their effort or time in defending their existing practices as opposed to improving them. To avoid this pitfall, it is important to engage the clinicians in selecting the performance indicators and the severity index. Managers can ask clinicians what they want to be evaluated on and how they wish to measure the severity of their patients. Furthermore, it is important to make sure that feedback to each clinician is provided privately and without revealing the performance of any single peer provider. It is okay to share the average of the peer providers, as long as the identity of each provider remains anonymous. The focus of feedback should be on everyone, not just on the clinicians with poor performance. An environment needs to be created where no one is blamed and all clinicians are encouraged to seek improvements as opposed to argue about the results.
3. *Inadequate measure of severity may mislead the analysis.* A poor severity index, one that is not predictive of the patients's prognosis, might give the impression of severity adjustment but in reality be no better than a random guess of outcomes. In these circumstances, an unadjusted benchmark is better because at least it does not give the appearance of what it is not. To avoid this pitfall, it is important to select a severity index that has high predictive power.
4. *Too much measurement may lead to too little improvement.* Sometimes analysts who conduct benchmark studies take considerable time to collect information and analyze it. In these circumstances, too little time may be spent on discussing the results, selecting a new course of action, and following up to make sure that the change is an improvement. It is important to keep in mind that the goal of benchmarking is improvement. Conducting an accurate analysis is only helpful if it leads to change and improvement; otherwise, it is a waste of time.

For more details about the risk of benchmarking, see Iezzoni 1997; Hofer et al. 1999; and Krumholz et al. 2002.

Presentation of Benchmarked Data

Presenting benchmarked data should be done in a fashion that helps clinicians improve their practices as opposed to act defensively. Poorly presented information leads to unnecessary and never-ending debates about the accuracy of the information presented. The Agency for Healthcare Research and Quality, the Centers for Medicare and Medicaid Services, and the Office of Personnel Management sponsored a working group on how to talk about healthcare quality.[1] The group made a number of suggestions on the presentation of benchmarked information. The following is a modification of the group's suggestion to fit within the context of benchmarking clinicians.

Before the Meeting

Check the Data

A simple mistake in benchmarked data will undermine the perception of the validity of the entire data set. To avoid this, check the accuracy of the data thoroughly before the meeting, making sure that all variables are within range and that missing values are appropriately handled. Prepare histograms of each individual variable in the analysis and review them to make sure they seem reasonable.

Prepare Graphs and Pictures

To help clinicians have an intuitive understanding of statistics, it is important to provide them with visual displays of data. Show data and summary statistics using bar charts and x-y plots.

Plan to Do it in Person

It is important to present benchmarked information in person to clinicians, allowing open question-and-answer periods. The presentation session should be scheduled ahead of time and well in advance.

Prepare handouts for discussion during the session. Distribute handouts to participants ahead of the meeting. Make sure that handouts are stamped "draft" and that the date of final report is clearly reported.

Prepare Stories

Supplement numeric data with anecdotal information that conveys the same message. Make sure that the anecdotes do not reflect judgments about quality of care but focus on the data being reported (i.e., patient's condition or patient outcomes). Provide an example of a typical patient complaint (usually in the form of a short video- or audiotape). It is important to weave the story or the anecdotal data with the voice of the customer.

At the Meeting

Confidential Evaluation

Make it clear that the evaluation is confidential. If you are talking to a group of clinicians, do not identify who is the best or worst. The analyst may let each clinician privately know how they performed against the group, but should not provide this information publicly.

Brief Introduction

Make a brief introduction of the purpose of the session. Introduce your project team, and ask clinicians to introduce themselves. Even if they know each other, still ask them to introduce themselves so they feel more comfortable in the meeting setting.

Limitations of Benchmarking

Acknowledge the limitation of the practice profiling method. Explicitly say that numbers could be misleading if the measures of severity are not adequate or the sample size is small. Point out that the focus should be on improvement and not measurement issues.

Start with a Customer's Story and Voice

Start the meeting by playing a brief tape of a customer talking in her own words about what happened to him. Use both positive and negative vignettes.

Present the Data and Not the Conclusions

Present the findings without elaboration about causes or explanations. Do not give advice about how clinicians can do things differently. Say, for example, "Data show that patients stay longer in our hospital than in comparable hospital" rather than "You should shorten the time it takes to discharge hip fracture patients." It is up to the clinicians to change and decide how to change; benchmarking just points out the variation in outcomes and facilitates the clinicians to focus on specific issues. What clinicians do depends on them. During the presentation, the analyst guides the clinicians through the data. The data and not the analyst help clinicians to arrive at a conclusion and to act.

Ask for Input

Explicitly ask for the audience's evaluation of the data after each section of the report is presented. Allow the clinicians to talk about the data by pausing and staying quiet. Say, for example, "Data show large variations on how long it takes us to discharge a patient with hip fracture. What do you think about that?" Pause and let participants talk about the variation in hip fracture data. The point is not to troubleshoot and come up with solutions on the spot but to discuss the issues and think more about causes.

Accept Criticism

The analyst should not defend the practice profiling method, the benchmarking effort, or any aspect of the analysis. Let the work speak for itself and accep suggestions for future improvements. Shift the discussion from blaming the study to what can be done to improve in the future.

Plan for the Next Step

Thank the clinicians for their time and describe next steps (e.g., "I will correct the report and get it back to you within a week.").

After the Meeting

Revise the Report

After the meeting is complete, the analyst should summarize the comments made during the meeting and append it to the report. Also, the analyst should describe the resources that were available to the clinicians (e.g., travel funds that were provided to attend meetings for a presentation of best practices).

Distribute the Report

Send a written report to each clinician. Please note that reports have a way of lasting well beyond the time for which they were generated. Make sure that all providers' identities are removed.

Ask for Feedback

Ask the clinicians to comment on the following:

1. What worked well regarding the practice profiles, and what needed improvement?
2. Do clinicians plan to change their practice? If so, in what way?
3. Was it worthwhile to gather data and do benchmarking? Why or why not?

Announce the Next Report

Once the reports have been completed and distributed and after the analyst has asked for feedback, it is time to schedule the next round of benchmarking.

Summary

This chapter outlines the use of decision analysis for benchmarking clinicians. Clinicians associated with poor quality of care negatively impact the organization by decreasing productivity and market share. Clinicians can be evaluated by examining patient outcomes, such as morbidity or client satisfaction, so that managers can compare the performance of clinicians without infringing upon the doctor-patient relationship. Several methods for benchmarking clinicians are described in this chapter, including benchmarking with and without risk adjustment. Benchmarking without risk adjustment compares clinicians to the average performance of peers. Benchmarking with risk adjustment compares clinicians to peers based on patients of similar severity of illness; when severity of illness is not known, patients of two clinicians can be matched feature by feature, creating an event tree. The use of an event tree allows an analyst to evaluate what will happen if one physician would have taken care of another clinician's patients.

Review What You Know

In the following questions, assume that you have collected data for two clinicians, Smith and Jones, and constructed the decision trees in Figure 12.6:

1. What is the expected length of stay for patients of each of the clinicians?
2. What is the expected cost for Dr. Smith if he were to take care of patients of Dr. Jones?
3. What is the expected cost for Dr. Jones if she were to take of patients of Dr. Smith?

Rapid-Analysis Exercises

Through Medline, select a disease area where a decision analysis of preferred course of action is available. Suppose you would like to benchmark three clinicians who are practicing in this area. Using the analysis found, carry out the following activities:

1. Design the forms you would use to collect information about patient's severity of illness (base the severity index on a measure published in the literature) and patient outcomes (use a standardized instrument).
2. Complete the forms for ten hypothetical patients using arbitrary answers.
3. Create a report analyzing the data.
4. Prepare for the presentation of your data (anecdotes telling customer's story, graphs, text of the presentation) using the procedures outlined in the chapter.

Audio/Visual Chapter Aids

To help you understand the concepts of benchmarking clinicians, visit this book's companion web site at ache.org/DecisionAnalysis, go to Chapter 12, and view the audio/visual chapter aids.

Note

1. See http://www.talkingquality.gov/docs/section1/default.htm.

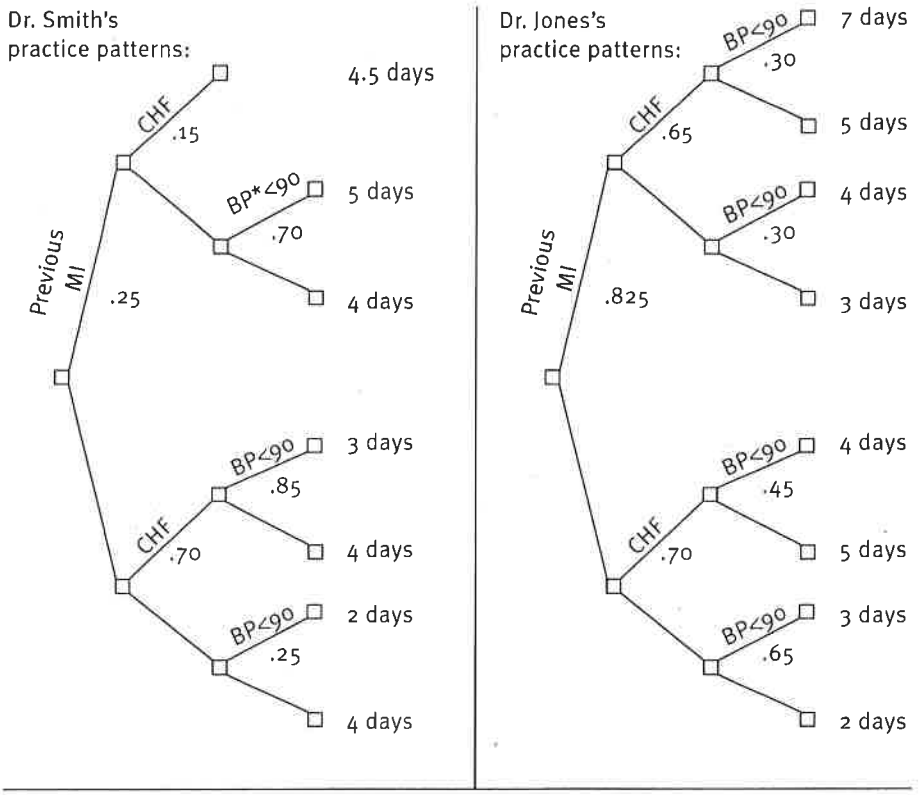

FIGURE 12.6 Practice Patterns of Two Doctors

* BP = blood pressure.

References

Alemi, F., M. R. Haack, and S. Nemes. 2001. "Continuous Improvement Evaluation: A Framework for Multisite Evaluation Studies." *Journal of Healthcare Quality* 23 (3): 26–33.

Callahan, M., O. Fein, and D. Battleman. 2002. "A Practice-Profiling System for Residents." *Academic Medicine* 77 (1): 34–9.

Hofer, T. P., R. A. Hayward, S. Greenfield, E. H. Wagner, S. H. Kaplan, and W. G. Manning. 1999. "The Unreliability of Individual Physician 'Report Cards' for Assessing the Costs and Quality of Care of a Chronic Disease." *JAMA* 281 (22): 2098–105.

Iezzoni, L. I. 1997. "The Risks of Risk Adjustment." *JAMA* 278 (19): 1600–7.

Krumholz, H. M., S. S. Rathore, J. Chen, Y. Wang, and M. J. Radford. 2002. "Evaluation of a Consumer-Oriented Internet Health Care Report Card: The Risk of Quality Ratings Based on Mortality Data." *JAMA* 287 (10): 1277–87.

Tversky, A. 1977. "Features of Similarity." *Psychological Review* 84 (4): 327–52.