Safety Performance Measurement:

IDENTIFYING PROSPECTIVE INDICATORS WITH HIGH VALIDITY

By R. SCOTT STRICOFF

Measurement is a current "hot" issue in safety. Safety professionals, union leaders, regulators and managers are all dissatisfied with the status quo—reliance (almost exclusively) on recordable and lost-time injury rates as safety performance measures.

The wisdom of seeking "upstream" measures has been recognized since total quality management concepts swept the U.S. more than a decade ago. In the field of quality management, Deming, Juran and other quality pioneers offered practical leadership regarding which quality factors should be measured upstream; they then provided consensus about the worth of doing so.

In safety, however, little guidance is available on how to move upstream. For example, many recent articles on TQM and safety sound like wishful thinking, as though the authors are saying, "Well, so much overlap exists between safety and quality, there must be some way to do safety like quality."

This article offers a survey of upstream and downstream measures currently available to safety practitioners and evaluates these indicators in terms of whether they are prospective or retrospective, and whether they have high or low validity as indicators (Figure 1).

TWO TYPES OF INDICATORS

Two types of safety measures are common in industry: accountability measures and performance indicators.

Accountability Measures

Accountability measures are a means of motivating people. They relate to specific performance expectations and specific people. The concept is not new; accountability through "management by objectives"-type systems has been common practice in industry for several decades. Petersen discusses these measures in some detail in his article "What Measures Should We Use, and Why?" and book *Techniques of Safety Management*.

Although accountability measures are important, one must not equate them with indicator or outcome measures. Activities are rarely direct predictors of results. This is not to minimize their importance as a management tool, but reliance on activity measures alone is analogous to grading a student on effort alone, with no consideration of actual achievement.

People who consistently and diligently perform an incomplete or incorrect set of tasks may nonetheless receive high marks for accountability—even though they have not produced desired outcomes. For this reason, Deming did not simply advocate accountability as the key to improving quality. Nor is it a simple answer for safety. Just as production quality requires a company to understand and measure upstream factors that permit intervention well before a defect occurs, safety management requires good upstream measures of the results that safety systems are delivering.

Performance Indicators

These measures indicate how successfully programs are achieving their ultimate objective—fewer injuries in the case of safety. Determining whether performance is improving requires a measurement system. In addition, a company may wish to benchmark: How do we compare to others in the industry? To other company plants or divisions? How does department A compare to department B within a plant?

An indicator may be characterized in terms of its timeliness and its validity vis-a-vis the outcome for which it is an indicator. Timeliness refers to when the indicator is measured (and to when it is measurable) in relation to the endpoint target. In safety, an indicator can be either prospective or retrospective (measured before or after the incident).

Validity refers to how accurately the measure predicts the endpoint event. A less-valid measure may provide a loose approximation of this event, while a highly valid measure reflects an actual count (Figure 2). Prospective measures are generally more useful yet have less validity than retrospective measures. Thus, the goal is to identify prospective measures that have the best possible validity.

To clarify these concepts, consider measures used to assess the financial health of a business. Profit is a highly valid downstream measure. It is counted with precision after the fact. Earnings estimates are less-valid upstream measures. They are prospective, but inherently have a degree of uncertainty (Figure 3). Wall Street expects these estimates to be highly valid; consequently, it is not unusual for stock prices to fall if actual earnings fall short of estimates by as little as one cent per share.

In safety, the most useful indicators are those that are prospective and valid (Figure 4). Such indicators provide information while there is still time to intervene and offer a high degree of confidence that intervention is occurring at appropriate points. In safety, advice is often readily available on accountability measures, but guidance on indicator measures is quite vague. This is likely because accountability measures are easy-to-implement—they reflect practices that have long been used in general management.

THE INJURY RATE AS A MEASURE

The safety profession has historically relied on one retrospective indicator: the injury rate. This is typically calculated using the following formula:

$$\frac{\text{number of incidents} \times 200{,}000}{\text{total hours worked}}$$

The standardized unit of measurement in this calculation is a workgroup of 100 people, each of whom works an average 40-hour week for 50 weeks each year; the result is a retrospective incident rate for the unit of 200,000 person-hours of exposure to workplace hazards.

This calculation is often the source of misinterpretation, however. For example, when a period of time passes with no injuries, some managers may interpret that to mean "good safety performance"; conversely, when several injuries occur within a short period, the typical response is, "Safety performance has deteriorated."

Either conclusion can be erroneous since a stable safety system will produce a variable number of injury events. To understand the significance of injury rate changes, one must statistically compare these outcomes with the predicted random outcome for the facility in question.

For example, suppose a workgroup consists of 100 employees. Its injury rate can easily change from 2.0 over a 12-month period to 4.0 for the most recent quarter despite no real change in the safety system. The prior 12-month period reflects two injuries, while the most recent quarter reflects one. Given the workgroup size, the quarterly increase to 4.0 is not significant (special-cause variation).

In safety, however, reward systems and performance appraisals are often based on numerical goals and measures that are untested for statistical significance. As a result, this group's supervisors may receive an undeserved bad performance rating.

Similarly, this workgroup could receive an undeserved good rating if the injury rate change is not statistically interpreted. If no injuries had occurred in the first quarter of the new year rather than one injury, the rate would be zero—a seeming improvement. In fact, however, this merely reflects random variation, not any change in the safety system.

How many workhours are needed for statistical validity? This depends on a site's injury frequency rate and the amount of variability present. The higher the frequency rate, and the lower the variability, the fewer hours needed. Standard statistical process control techniques work well with injury data, and the usual statistical guidelines provide an accurate answer for questions of validity.

For the typical facility, these statistical methods mean the injury frequency rate is a measure that is accumulating validity as time passes. Yet, this rate is of no predictive value to safety management on a monthly or even quarterly basis, let alone a weekly or daily basis. On the contrary, given the misplaced trust that people accord to the frequency rate, these numbers not only do not help the safety effort, they hinder proactive safety management.

Reliance on such incomplete statistics gives rise to the accident cycle. When the recordable rate exceeds a facility's upper-limit perceived acceptability, management acts to drive the rate down. When the rate falls below that limit, attention to safety declines, and the recordable rate rises again. In this cycle, management action for improvement follows fluctuations in the injury frequency.

So, despite industry's reliance on the injury frequency rate, its use is limited by several factors.

1) The performance outcome measured is so far downstream (retrospective) that it is unsuited for proactive safety management efforts.

2) Furthermore, the greater the emphasis on this rate, the less valid it becomes, as people learn how to "make the numbers come out right." As a result, the injury rate loses a key trait—it moves from being a highly valid retrospective indicator to being a retrospective indicator with low validity—the weakest and least useful type (Figure 5).

3) Because injury frequency rate and interpretations of rate changes are susceptible to distortion, the rewards and punishments based on them come to be seen as capricious and unjust.

FIGURE 1

Validity \ Timing	Prospective	Retrospective
High		
Low		

FIGURE 2

Validity \ Timing	Prospective	Retrospective
Higher		Actual Count
Lower	Loose Approximation	

FIGURE 3

Validity \ Timing	Prospective	Retrospective
Higher		Profit
Lower	Earnings Estimates	

FIGURE 4

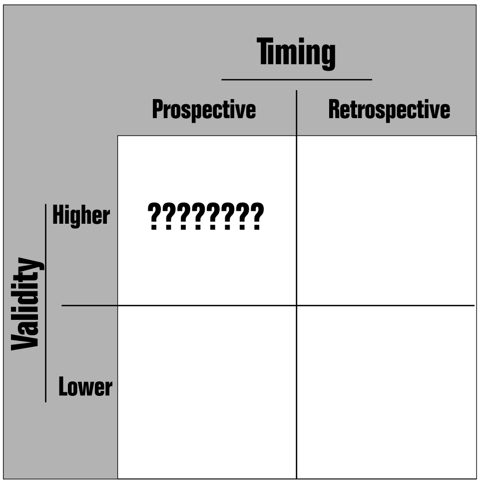


FIGURE 5

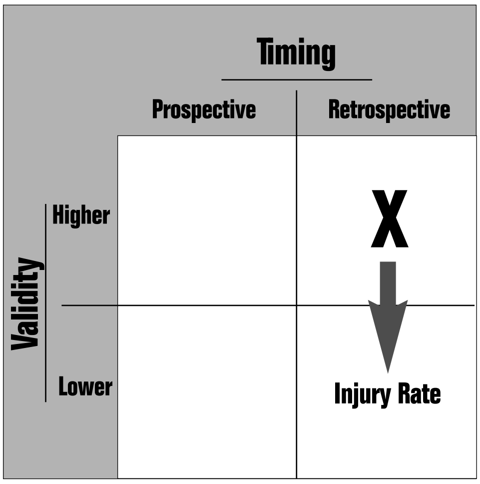
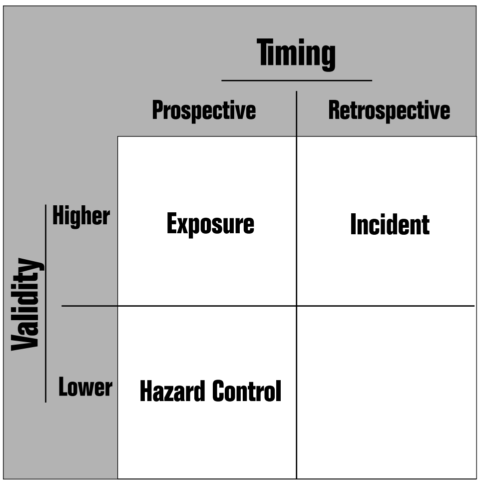


FIGURE 6



BABY WITH THE BATH WATER

In response to these issues, many within the safety community have advocated abandonment of the injury frequency rate. Despite the cited limitations, it would be both unrealistic and shortsighted to take such action.

This rate is the ultimate outcome measure for the safety process and is analogous to a business measuring profit. Although profit alone does not tell managers how to improve performance, accountability ultimately demands that the outcome be measured. Thus, the injury rate should not be abandoned, nor should it be accepted as the sole measure. Instead, safety managers must strive to supplement it with good (high validity) upstream (prospective) indicators.

MOVING TO UPSTREAM MEASURES

The first step in understanding effective upstream measures is to consider the "stream." Figure 6 offers a conceptual model of the pathway to injuries.

Hazards

The starting point is hazards; in the absence of hazards, no injuries will occur. The term hazard refers to something with the potential to cause injury (e.g., high-voltage electrical line entering a breaker box). Hazards are a fact of life. The only way to completely eliminate them is to eliminate activity. Firms can tally hazards as an upstream measure, but this is not meaningful as it ignores the impact of the many actions taken to control hazards.

Controls

Moving downstream, the next step is controls. These may be engineering, procedures or personal protective devices. Measurement of controls is generally the focus of compliance audits—a useful exercise, but not a good predictive measure. Audits evaluate the presence and design of control programs, with some retrospective evaluation of their implementation. However, audits do not effectively anticipate future implementation quality. In fact, at some sites, well-conducted audits may produce "clean" reports, while injuries continue unabated.

Many organizations gather data that is indicative of activity levels for hazard controls. For example, some track safety training or safety meeting attendance. Such metrics are not sufficient, however, unless their relationship to reduced exposure can be demonstrated. For a safety measurement system to be effective, a predictive linkage must exist between the parameter being measured and the outcome produced.

Exposure

Moving downstream again, one finds exposures. Exposures exist wherever hazards and controls are present since no control is perfect. This imperfection "opens the loop" between control activities and injuries and makes their linkage indirect—no matter how many controls are instituted, some injuries may still occur.

Design, maintenance and use of control equipment and procedures provide opportunities for breakdown as well—which result in workers being exposed to risk. Although not every exposure leads to injury, no injury occurs in the absence of exposure. Thus, if a firm measures exposures, it will have a good predictive (highly valid prospective) measure for injury.

At-Risk Behavior

Measuring at the exposure stage represents a move upstream from the occurrence of incidents. The advantage is that it uses a parameter which is directly proportional to injury incidence. In other words, while the relationship between control activities and injuries is indirect, the relationship between exposure and injury is direct. At-risk behaviors are indicative of exposure; thus, measuring them can be used to develop a good upstream indicator.

MEASURES FROM CONTROL PROGRAMS

Tracking activity in safety-related programs is a frequently cited example of an upstream measure. For example, a company may use safety meeting frequency and attendance, or number of job safety analyses completed as upstream measures for safety. One can see the logical appeal of basing an upstream measure on activities conducted to ensure and enhance safety. However, if not properly constructed, these measures can be misleading and counterproductive.

For activity tracking to be a valid upstream measure, a site must be able to demonstrate *through data* that a significant correlation exists between the frequency of the activity and the outcome it produces. Because safety performance is the result of the complex interaction among many factors, such correlations are difficult to find.

However, if a non-correlated measure is used as an indicator of safety performance, a firm may wind up diverting resources to activities that have little or no real impact. Furthermore, because the complex interaction that relates activities to outcomes differ at each site (based on factors such as production process, culture, worker experience, management style), the correlation should be demonstrable site-by-site in order to ensure measurement validity.

MEASURES FROM BEHAVIOR-BASED SAFETY

A behavior-based safety (BBS) process allows a company to track and report upstream parameters directly indicative of exposure. The validity (value) of these

In safety, the most useful indicators are those that are prospective and valid. Such indicators provide information while there is still time to intervene.

parameters depends on two factors: 1) design of the data-collection process; and 2) quality of its implementation.

In BBS processes that generate valid measurement data, correlations can be found between "percent safe" and injury rates. Similarly, correlations can likely be found between other parameters (such as observation rates, feedback rates, identification and removal of barriers to safe behavior, and observation quality) and subsequent injury rates. These become useful management tools—management is able to base interventions on likely outcomes and make decisions *before* performance degrades.

To produce "measurement quality" behavioral data, several key issues must be addressed.

Process Design

Four key elements must be considered in the design of a BBS process: identification of behaviors critical to exposure, operational definitions, sampling adequacy and decision rules.

When undertaking a BBS process, a site must select behaviors that will be observed. If the goal is to produce data that is predictive of performance, behaviors selected (which are actually exposures) must be representative of the exposures that produce injuries at a facility. If behaviors are selected based on perception or ease of observation, the end result will be data that does not correlate with injury performance.

To generate measurement quality data, behavior selection must be risk-based. This involves understanding what hazards exist and have caused—and are likely to cause—injuries.

In addition, critical behaviors must be well-defined. Observers must be able to consistently recognize and characterize behaviors in an objective manner. This only occurs when behaviors are operationally defined in unambiguous terms.

A third issue is sampling strategy—deciding when and where to observe. Two strategies are common in industry: 1) focus on "hot spots"; or 2) design representative sampling.

In the former approach, a disproportionate amount of observation effort is targeted at those tasks, departments and/or time periods that produce the largest number of injuries within the location. In the latter approach, observations are scheduled to cover the facility either uniformly or in proportion to the amount of work effort produced.

Since measurement quality observations can occur under either strategy, the key factor is consistency of the strategy chosen. However, the strategy selected does affect the sensitivity of observation data as a measure. With a "hot spot" strategy, data will be more sensitive to performance changes than will that gathered via a balanced or uniform observation strategy.

Decision rules are the fourth factor needed to produce high-validity measurement data. In this context, the term refers to conventions regarding the way behavioral definitions will be "scored" on the data-collection form. Although it may seem that unambiguous operational definitions eliminate the need for such rules, actual field applications show this is rarely true.

For example, decision rules are required to answer questions such as these. If two employees are working together and both use the wrong tool, is this counted as one at-risk behavior or two? If a worker uses an incorrect lifting technique to move three different boxes to the same dolly, does it count as one or three at-risk behaviors? If someone fails to remove a hazard (such as scrap in an aisle), does it count as an at-risk behavior? Without decision rules, data will be inconsistent, which will lead to misleading performance indicators.

Observer calibration is also a key factor. Observers must see the same things. That is, if two observers watch the same task, do they see the same behaviors occurring? If so, do they characterize each behavior the same way? That is, is the critical behavior's definition and the observers' training sufficient enough so that the same behavior will not be identified differently by two different observers? Do the observers have a shared understanding of what is safe and what is at-risk? Without calibration, data produced will be laden with artifacts that prevent its effective use as a predictive tool.

If a company desires BBS measures further upstream than the occurrence of at-risk behavior, it must consider the quality with which the intervention mechanism is used. Having a process in which upstream measures (such as observation rate) correlate with downstream injuries provides a powerful opportunity for intervention to occur before exposure. However, the active mechanisms of change must be effective for this to succeed. If measurement quality observation data is the goal, the process should have provisions for monitoring the quality of reinforcement provided and actions taken to remove barriers to safe behavior.

CONCLUSION

A thorough approach to safety measurement should encompass accountability measures, downstream performance measures and upstream performance measures predictive of outcomes. Managers who use safety performance measures must understand how to correctly interpret downstream measures; safety professionals must take the lead in ensuring that they develop this understanding.

As downstream safety measures are developed, measures at the control and exposure stage may be considered. However, at either point, one must ensure that the measures truly correlate with outcomes, and that the data-collection process produces valid, consistent information. Because BBS processes generate data at the exposure stage—directly upstream from injuries—they can be harnessed as a foundation for improved upstream measurement. ■

REFERENCES

Petersen, D. "What Measures Should We Use, and Why?" *Professional Safety.* Oct. 1998: 37-40.

Petersen, D. *Techniques of Safety Management.* New York: McGraw Hill, 1978.

R. Scott Stricoff*, CSP, CIH, is president of Behavioral Science Technology Inc., an industrial safety consulting firm headquartered in Ojai, CA. He is a professional member of ASSE's Valley Coastal Chapter, and a member of American Industrial Hygiene Assn. and American Academy of Industrial Hygiene.*