

A Proposal to Change the Office of Science  
Injury Reduction Contractual Performance Measures

By William J. Griffing

Abstract

All ten of the Office of Science (SC) national laboratories have made significant progress in transforming the safety culture of their organizations over the last five to 10 years. The gradual downward trend in injury rates at each of these laboratories is evidence of this transformation. The present practice of inserting uniform injury-reduction metrics into each SC laboratory's contract, despite our differences, and attempting to measure that performance with limited sampling data, however, does not result in accurate assessments of performance. This paper offers an alternative approach to performance measurement.

**All SC laboratories are not the same**

As safety directors we know from experience that it is not reasonable to expect that all operating organizations within our individual laboratories can attain the same level of injury-reduction performance. We know that our operations and maintenance mechanics, for example, encounter greater hazards in the course of their daily routines than do our computer professionals. Although there is equal pressure applied to both of these organizations to eliminate injuries, we understand (and can predict) that the injury rate for the former will most likely be higher than the latter. We would not conclude that they were doing "equally well" just because they had the same injury rates. What we expect of each of our line organizations is to continue to improve against its own history of performance, knowing that the cumulative effect will be an improved overall rate for the entire laboratory.

The Office of Science should apply the same logic in its expectations of injury-reduction performance to the 10 national science laboratories. Each laboratory's mission is different. The type and composition of work varies among the laboratories – sometimes dramatically. Certainly the composition and characteristics of the workforces at each of these laboratories differs significantly. The use of subcontractors varies, the amount of major construction occurring in any given year varies, and the degree to which individual state workers' compensation laws affect the injury metrics varies.

It should be the stated long-term objective of the Office of Science for its laboratories to establish and nurture strong safety cultures. Partial evidence of this will be, over time, continued reductions in the injury rates as each laboratory progresses along a path of continuous improvement. The Office of Science should look for evidence of this continuous improvement by comparing each laboratory's current performance against its own history of performance. The present scheme, where all laboratories are measured by the same yardstick does not yield a valid assessment of performance.

**The present problem**

As the frequency of DART case occurrences has declined at each of the ten SC laboratories, the accuracy of the DART rate measurement has also declined. This is due to the fact that the time period included in the calculation has remained constant at one year. The following will illustrate how the degree of accuracy of a computed injury rate declines with a reduction in the number of observations (injuries), and how it can be improved by extending the sampling period.

It should be acknowledged that a laboratory's injury rate is something that can't ever actually be measured. We can only approximate that number by counting the number of observed events (injuries) over some period of time. The rate we calculate will obviously be different for each period of time we choose to "sample." Clearly the larger the number of events within a given sampling period allows for a greater degree of confidence in the computed approximation of the "real" injury rate. (Remember, we can never know our real injury rate, only approximate it based on our sampling.) Consider the following:

The DART case rate is the number of such events recorded per 100 person-years of work. To try to identify the statistical errors associated with the measurement of such case rates, it must be recognized that DART cases correspond to discrete events; the number of such events during any given period of time used for measuring this rate will be an integer. A reasonable assumption one can make is that these events are random and have a normal (bell-curve) Gaussian distribution. In a well-managed workplace these events are very rare, so it is reasonable to apply Poisson statistics.

Thus, if one has  $n$  events during a specified period of time,  $t$ , the standard deviation in the number of events is

$$\sigma = \sqrt{n} .$$

However, what is of interest is the measured rate  $r = n/t$ , and its standard deviation  $\sigma_r$ ,

$$r \pm \sigma_r = \frac{n}{t} \pm \frac{\sqrt{n}}{t} ,$$

where we have assumed that the time period of the measurement,  $t$ , has no "error" associated with it. Proceeding a bit farther:

$$\sigma_r = \frac{\sqrt{n}}{t} = \sqrt{\frac{n}{t} \frac{1}{t}} = \sqrt{\frac{r}{t}} .$$

Finally, the relative error of the measurement of the DART case rate is  $\frac{\sigma_r}{r}$ .

First, consider a large laboratory with 2000 workers. For such a laboratory, a DART case rate of 0.5 corresponds to a rate,  $r$ , of 10 events per year. In the following table, one can compare the values of the various quantities for three different selections of the measurement period, 0.25 year, one year, and three years:

$t$	3 months	1 year	3 years
$r$	10	10	10
$\sigma_r$	$\sqrt{\frac{10}{0.25}} = 6.32$	$\sqrt{\frac{10}{1}} = 3.16$	$\sqrt{\frac{10}{3}} = 1.82$
$\frac{\sigma_r}{r}$	0.632	0.316	0.182

One can see that the 3-month measurement period is absurdly short given the large relative error. The standard deviation of the DART case rate will be the DART case rate multiplied by the relative error shown in the bottom row of the table. Thus, for the 3 month measurement period, the DART case rate, with one standard deviation errors, is  $0.5 \pm (0.632 \times 0.5) = 0.5 \pm 0.316$ . The consequence of these rather large relative errors is that one cannot statistically distinguish between DART case rates ranging from 0.184 to 0.816 with the level of statistical confidence associated with a one standard deviation measurement. Similarly, for the one-year measurement period, one cannot statistically distinguish between DART case rates ranging from 0.342 to 0.658.

Let's see how this works for a laboratory with a worker population of 1000. For this laboratory, the DART case rate of 0.5 corresponds to 5 cases per year. Filling out the same table, we have:

$t$	3 months	1 year	3 years
$r$	5	5	5
$\sigma_r$	$\sqrt{\frac{5}{0.25}} = 4.47$	$\sqrt{\frac{5}{1}} = 2.24$	$\sqrt{\frac{5}{3}} = 1.29$
$\frac{\sigma_r}{r}$	0.894	0.447	0.258

It is easy to see that the relative errors are very large for the 3-month and 1-year measurement periods, and even considerable for the 3-year time window.

Now consider a really small laboratory with only 200 employees. For this laboratory, the DART case rate of 0.5 corresponds to one case per year. Filling out the table once again, we have:

$t$	3 months	1 year	3 years
$r$	1	1	1
$\sigma_r$	$\sqrt{\frac{1}{0.25}} = 2$	$\sqrt{\frac{1}{1}} = 1$	$\sqrt{\frac{1}{3}} = 0.58$
$\frac{\sigma_r}{r}$	2	1	0.58

For all three measurement periods, the relative errors are large and could be said to be meaningless for the 3-month and 1-year measurement period.

In conclusion, with DART case rates in the neighborhood of 0.5, the statistics of rare events incorporate relative errors that limit the ability to make valid comparisons. Even for the 2000

worker laboratory and a three-year measurement period, the DART case rate can only be measured to an accuracy of about 20 per cent. For smaller laboratories, the accuracy of the measurement is dramatically worse. This is yet another reason why comparisons between laboratories with significantly different size worker populations is inadvisable.

### **A path forward**

Beginning with FY06 contract negotiation discussions, it is suggested that the Office of Science issue new guidance to site offices for negotiating injury-reduction goals with their respective laboratories. This guidance should emphasize that the goal of such measures is to demonstrate a path of continuous improvement when compared to each laboratory's own history of performance and to no one else's. Improvements in the range of five to ten percent per year are considered realistic gains.

Further, the Office of Science guidance should also point out that it would be advisable to negotiate injury-reduction rates that are calculated using a three-year period (the "rolling average method") to decrease the error in the measurement due to the small number of events observed within the span of a single year's performance period. This should be done for both the DART case rate as well as the TRC rate. Doing so will allow for meaningful measurements of improvement – something that is not now possible.

### **Special Thanks**

I am grateful to Don Cossairt for assisting me with the statistical calculations used in this paper.