PERGAMON

Safety Science 32 (1999) 1–18

SAFETY SCIENCE

Long-term evaluation of a behavior-based method for improving safety performance: a meta-analysis of 73 interrupted time-series replications

T.R. Krause*, K.J. Seymour, K.C.M. Sloat

Behavioral Science Technology, 417 Bryant Circle, Ojai, CA 93023, USA

Accepted 22 December 1998

Abstract

Research and applications of behavioral principles have established behavior-based safety initiatives as potentially effective solutions to certain occupational health and safety challenges. The present study adds to the existing literature a longitudinal evaluation of an employee-driven behavior-based accident prevention initiative implemented in industrial settings. Up to 5 years of injury data from 73 companies, drawn from a target population of 229 companies who implemented behavior-based safety, were examined. Comparisons of pre- to post-initiative incident levels across groups revealed a significant decrease in incidents following the behavior-based safety implementation. Effect sizes were estimated from the average percentage reduction from baseline. The average reduction from baseline amounted to 26% in the first year increasing to 69% by the fifth. These findings are critically examined in terms of both internal and external validity. Future research will focus on differential effects of specific elements of the behavior-based safety initiative described herein. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Behavior-based; Industrial safety; Accident prevention process

1. Introduction

Too often, the researcher and practitioner work in isolation from one another, in spite of the value of blending their roles. The practitioner often comes across valuable knowledge useful to the researcher, yet it is rare that a consultant finds a way to

*Corresponding author. Tel.: +1-805-646-0166; fax: +1-805-646-0328; e-mail:trkrause@bscitech.com

PII: S0925-7535(99)00007-7

conduct research related to their own services. Non-research directed missions, difficulties with data collection in the field, and the pressure to make business decisions before long-term research become available all discourage practitioners from conducting outcome research. The present study was a safety consulting organization's effort to systematically evaluate the results of its clients.[1]

Behavioral approaches to safety performance improvement have enjoyed increasing recognition as effective solutions to occupational health and safety challenges. Some of the most frequently cited components of behavior-based safety are goal-setting and posted feedback (e.g. Chhokar and Wallin, 1984; Reber et al., 1984), observation and posted feedback (e.g. Komaki et al., 1980), and observation, verbal feedback, data analysis, and problem solving (e.g. Krause et al., 1990; Krause, 1995). McAfee and Winn (1989) described the content of various behavior modification programs in greater detail and summarized 24 published studies. Guastello (1993) conducted a meta-analysis of seven studies on the effectiveness of behavior-based safety programs. The common features of these programs were employee training, observation, and feedback. However, they differed in their use of goal-setting, incentives, method of observation, type and method of feedback, and in their longevity. In the present approach, front line employees of the implementing organization were trained to conduct observations, provide feedback, analyze behavioral data, and use behavioral data to make system improvements.

2. Background of employee-driven, behavior-based safety

Komaki et al. (1978) were among the first researchers to apply behavioral techniques to industrial safety when they worked with a food manufacturing plant. In their study, operational definitions of safety-related behavior were developed and used to observe and record behavior on a checklist. The experimenters provided posted and verbal feedback to workers 3–4 times weekly, based on data gathered during observations. Results showed improvements in safe behavior of 26% in one department and 21% in another. When the observation program was withdrawn, performance returned to baseline. This 25-week project did not examine the effects of the initiative on injury rates.

The pioneering research of Komaki et al. (1978) demonstrated that applied behavior analysis methods could be used effectively to improve safety-related behavior in the workplace and, at the same time, raised several questions. If safety-related behaviors improved, would comparable change in incident rates follow? How could organizations learn to implement and maintain behavioral methods themselves without dependence on outside consultants? How can site personnel sustain performance improvement? Who should conduct observations and give feedback? What types of feedback work best? To what degree would findings of concomitant changes

[1] While the specific technology described and evaluated in this study is owned by a company for whom its authors work, every effort has been made to conduct and report this evaluation without bias. Nevertheless, readers should be aware of this fact when assessing the findings.

in incident frequency and safe behavior generalize to other behaviors not directly measured?

Subsequent research has addressed some of these questions and neglected others. That specific knowledge of results is important was shown by Locke et al. (1981) and that specific feedback facilitates development of a more detailed plan for behavior change (Kopelman, 1986). Kopelman also reported that when feedback revealed a performance deficit, it was important to develop an action plan to take corrective action. Earley (1988) demonstrated that action planning mediated the influence of specific feedback. Specific feedback stimulated more detailed planning than general feedback and led to a higher level of performance.

While behavioral scientists learned about feedback and its relationship to goal-setting and action planning, organizational theorists were developing methods for managing quality. Deming's (1986) work on management systems and their effect on quality transferred directly to industrial safety management. Krause et al. (1990) point out that much of the observed variation in injury rates is attributable to random variation. They, as well as others in the field, recognized that when statistical process control charts indicate that incident rates are in control (i.e. most observed variation is random), system changes are called for if incident rates are expected to improve (Krause and Hidley, 1989; Salazar, 1989; Krause, 1995).

Deming (1986) held staunchly to the view that it is management's responsibility to change the system. Krause and Hidley (1989) combined this approach with applied behavior analysis and targeted safety as an improvement area. Management controls training resources, develops and implements policies and procedures, regulates spending for equipment, and selects and places personnel. When management understands its responsibility for employee safety and directs improvement efforts on the safety system it created, then management also understands that blaming the employee will not result in safety improvements. Deming (in quality) and Krause and Hidley (in safety) prescribed adequate training, measurement of upstream process indicators, in-depth assessments, and positive feedback on the status of the system. At the same time, involvement of employees at all levels was recognized as the mechanism for continuous improvement in safety.

Taken together, these developments in organizational behavior management, combined with methods from total quality, suggested some important possibilities for industrial accident prevention. Krause (1989, 1992) confirmed Komaki et al.'s (1978) finding that the most effective observation strategy is to target site-specific behaviors that have led to incidents in the past. Rather than using a standard list of behaviors that may or may not be applicable in a particular setting, Krause and colleagues adapted quality improvement techniques into standard methods for identifying relevant behaviors from incident reports (Krause et al., 1984, 1990; Killimett, 1991a). They developed a method for providing immediate positive verbal feedback following observation (Krause et al., 1990; Killimett, 1992; Krause, 1997).

Krause and his colleagues involved employees at all levels of the organization in order to develop a self-sustaining mechanism for continuous improvement. This approach replaces the traditional model of training supervisors and managers to observe behavior and deliver feedback, or having outsiders perform this function.

Employees across levels learned to observe and give feedback, and then to use behavioral data to select safety improvement targets.

Managers and supervisors play an important role in removing barriers to safe performance and facilitating the smooth operation of the process. Because employees implemented the safety initiative, they possessed the tools to train new participants and were personally invested in the improvement effort (Killimett, 1991b; Krause, 1995). This approach defines the phrase 'employee-driven, behavior-based safety process'.

Most studies of the effectiveness of behavior-based safety programs, including those cited by McAfee and Winn (1989) and Guastello (1993), have been predominately time-limited, lasting weeks and months rather than years. These studies rely heavily on external resources for implementation and did not provide a sustaining mechanism.

To date, a systematic and long-term evaluation of the effect of behavior-based safety with observation, feedback, data analysis, and action planning components has not been reported. The present study is the first in a series of process and outcome evaluations. It adds to the existing literature a longitudinal evaluation of an employee-driven accident prevention process implemented in real-world industrial settings.

3. Materials and methods

3.1. Participants

Participants were 73 facilities who responded to requests for data, from a population of 229 facilities who met the inclusion criteria for this study. Three inclusion criteria reduced the 407 sites using the employee-driven behavioral accident prevention down to the target population of 229. First, each steering committee must have been trained by an outside consultant, as opposed to having been trained by an employee of their company (internal consultant). The internal consultant model of implementation uses a different implementation strategy than the one being investigated here and is the subject of a separate study. Each site also had to be located in North America and had to have completed at least 1 year of observations. This long-term independent variable was necessary to allow adequate time to affect the global and long-term outcome measure, incident rate.

All sites were asked to send monthly, non-cumulative recordable cases and hours worked, covering only those employees potentially affected by the behavioral process, beginning 4 years before behavioral observations began. To give an idea of how much data was being requested, a site beginning observations in 1987 would have been asked to compile as many as 336 numbers from as many separate records. Considering that it was typical for a site never to have compiled their monthly injury data, this was a substantial undertaking. Moreover, for organizations not previously inclined to gather and analyze monthly incident data, the task of motivating a response for this project was even more difficult. Sites gave the following types of reasons for not responding: "The data are stored in another location and I don't have access to it", "I don't know where to find that data", "It would take too much time to pull all that together", "I only have summary reports", "I can't get the monthly hours worked from the personnel department", "We changed our recording criteria and didn't apply the new criteria retroactively", and "We've had a major reorganization, so it's impossible to get injury data for the group of employees doing the behavior-based safety process".

Because only 73 of the 229 target sites provided data, selection bias was a concern. To better understand the extent of selection bias, the 73 participants were compared to the 156 non-participants on several variables: longevity of the process, number of employees, union representation, industry, and responses to the question "Do you view your (behavior-based safety) process as a success?"[2] Means and percentages of these variables for each group are shown in Table 1. Independent groups *t*-tests assuming unequal variances revealed that the participants extended their behavior-based safety process to more employees ($M = 533$) on the average than non-participants ($M = 363$), and that this difference was significant ($t(110) = 2.60$, $p < 0.01$). This finding suggests that larger facilities may have been able to devote resources to data gathering. Similarly, participants had newer processes, or fewer years of observation ($M = 3.11$ years) than non-participants ($M = 4.15$ years). This difference in longevity was also significant ($t(122) = 3.56$, $p < 0.01$). A chi-square test showed that the distribution of sites across industries did not differ between participants and non-participants, and binomial tests showed that the groups did not differ in being represented by a union or not. Finally, the groups did not differ in response to the question "Do you view your process as a success?" This is an important

Table 1
Comparison of participants to non-participants

	Participants		Non-participants	
	M or %	*n*	*M* or %	*n*
Number of employees participating*	533	73	363	156
Longevity (years since observations began)*	3.11	73	4.15	156
Union representation	48%	73	57%	156
Industry		73		156
Chemical	30%	22	27%	42
Paper	21%	15	15%	23
Petroleum	12%	9	15%	23
Other	37%	27	43%	68
"Do you view your process a success?" (Yes)	97%	58	91%	92

*Two-tailed $p < 0.01$.

[2] As part of a separate and unpublished research project, the employees responsible for coordinating the behavior-based safety processes (the facilitators) at 150 sites who were also members of the target population in the present study were interviewed. One of the questions was "Do you view your process as a success?"

non-indication of selection bias: if strong selection bias were operative we would have expected to find differences in response to this question.

Because of significant differences between participants and non-participants on the two variables, size and longevity of process, post hoc analyses were conducted to determine the likely impact of these differences on the effect size estimates.

4. Procedure

4.1. Design

This study utilized an interrupted time-series design with 73 replications. A meta-analysis combined results at individual facilities for the larger study. At each location studied, the independent variable was the behavior-based safety initiative, described in the following section, and the dependent variables were the United States of America Occupational Safety and Health Administration (OSHA) recordable injury/illness rates. In the US, OSHA recordable cases are "All work-related deaths and illnesses, and those work-related injuries which result in: Loss of consciousness, restriction of work or motion, transfer to another job, or require medical treatment beyond first aid" (US Department of Labor, 1986, p. 17). Recordable rates were computed by multiplying the number of recordable injuries in a given period by 200,000 and dividing by the number of hours worked in the same period. This is roughly equivalent to the number of recordable cases per 100 employees per year.

4.2. Independent variable

All 73 facilities received training for a behavioral science approach to accident prevention originally described in Krause et al. (1990) and later in Krause (1997). Behavior-based safety reduces incidents through management of at-risk behaviors. This approach, which identifies, measures, and improves safety-related behaviors, is based on measurement, upstream sampling of key variables, feedback, problem solving, and employee involvement. It identifies and corrects existing systems that produce at-risk behavior and develops new systems that encourage safe behavior.

Although each behavior-based safety initiative was unique, they shared several important features. At each site, a safety assessment was conducted. The consultant evaluated existing strengths and weaknesses via direct observation, interviews, surveys, and analysis of incident data. The consultant's written assessment report to the organization included a recommendation for the implementation plan.

Following each company's assessment, at least one steering committee was formed. The steering committee would be trained to become the driving force behind the initiative. This group acquires the knowledge and develops the skills and experience required to implement behavior-based safety. This built-in mechanism kept the processes functioning over the long-term without constant input from an outside consultant. Committee membership was approximately 80%

non-management employees and 20% first-level supervisors. The chairpersons were usually non-management employees.

Each committee's first tasks were to develop a checklist of critical behaviors and write operational definitions for each behavior. Each implementation was unique in that the number and definitions of critical behaviors varied depending on the behaviors that had contributed to incidents in the past. For example, two sites identified "Use of Barriers and Guards" as critical behaviors. The steering committee at one chemical company defined it as: "Barricades are placed around all jobs where entry by an unsuspecting person could result in an injury", e.g. "Barricades shall be placed just outside the swing radius of TEX 554, 555, and 557". The steering committee at an electric company developed a slightly different definition: "Is the area/equipment that is at risk properly identified to make people aware of the danger?" For example: "When grating is removed, does the person install red barricade tape? Does person install "Danger High Voltage" sign when leaving cubicle open with voltage present?" Each site studied had developed their critical behaviors checklist and definitions from an analysis of their incident reports whereby they extracted behaviors that had contributed to incidents in the past and defined them using examples that made sense for their unique situation.

Once the checklist was developed, the steering committees selected and trained observers to use the checklist to observe behavior and then to give feedback. When presenting the feedback, the observers learned to begin by describing the observed behaviors which were completed safely (success feedback). Next, they learned to describe the at-risk behaviors they observed, and discuss these with the employee(s) (guidance feedback). They learned to use this discussion in a positive way to discover exactly what happened that was at-risk and why. Typically, observer training lasted 2 days. It included education on the foundations of behavior-based safety, information on how the critical behaviors checklist and definitions were developed, classroom training on how to identify at-risk behavior (usually slides depicting safe and at-risk behaviors), classroom training on interaction skills, and field training. During the field training, new observers would conduct observations under the guidance of a steering committee member who served as a coach.

Finally, the steering committees learned to use observation data to identify and solve systemic problems. After each observation, the scores from the data sheets were entered into a database anonymously (so as not to identify individual employees). Typically, steering committees printed monthly reports showing, for example, the percent safe score for each behavior. Employees then had the opportunity to select high-risk behaviors for problem solving.

On the average, the steering committee training described above took place over 9 months for each site. Because feedback is one of the key mechanisms to improve safety performance in a behavior-based safety initiative, the date observations began was used as the start date for each project. Observations typically began in the 6th month of implementation. In the present study, baseline was defined as the 4-year period prior to the date observations began, while the follow-up period began immediately after observations began and extended through to the present.

4.3. Materials

Each steering committee received training manuals, slides, and a software package for data collection.

4.4. Data collection procedures

In June 1993, a task force was formed for the purpose of data collection on this project. The effort began with a letter requesting monthly non-cumulative US OSHA recordable rates beginning 4 years before implementation (hereafter referred to as 'results data'), in exchange for a statistical analysis. Although the same results data were requested from each organization and every effort was made to gather the complete results set (multiple phone calls, providing forms, discovering why data were not sent and removing barriers wherever possible), sites were encouraged to submit whatever data they could retrieve from their records. Beginning in 1996, all target organizations were also offered complimentary participation in a behavior-based safety benchmarking service[3] in exchange for the anonymous use of the data for research.

Of the 229 sites that met the inclusion criteria described above, 120 provided OSHA recordable injury/illness rates[4]: a 52% response rate. Of these, 47 were excluded because they failed to provide enough data to permit statistical tests. Table 2 provides details on the inadequacies of these data. It shows the minimum requirements for monthly and annual data, and the frequency with which sites failed to provide the minimum by period (baseline, follow-up, or both).[5] This left 73 sites, or 32% of the target population, to include in the study. Nineteen of the 73 participants had sent annual data, four sent quarterly, while 50 sites sent monthly US OSHA recordable rates. Thirty-three of the participants shared their data through the benchmarking service, while the remaining 40 participants shared their data for other reasons (e.g. response to authors' request, custom analysis).

Table 2
Frequency of failures to send adequate data by type of data and period

		Period		
Type of data	Minimum requirement	Baseline	Follow-up	Both
Monthly	12 months per period	2	6	28
Annual	2 years per period	2	0	9

[3] This is a complimentary service available to the entire population of interest to this study. Sites use the benchmarking service, which produces quarterly benchmarking reports, to compare process and outcome measures, to identify best practices in behavior-based safety, and to network with one another.

[4] Two sites provided their best estimates of the OSHA recordable injury/illness rates. Of these, one site did not report to OSHA but provided a near equivalent based on medical treatment, lost time, and restricted duty cases. The other site provided recordable rates based on disabling injuries/illnesses alone.

[5] Adequate data are continuously solicited from sites and incorporated into the meta-analysis. The most current results are available from the authors.

5. Results

Prior to analysis, each data set was entered and verified for accuracy of data entry. Data were also screened for inconsistencies. For example, if a site provided both annual and monthly injury counts, the monthly numbers were aggregated and the results compared to the annual figures. All discrepancies were reconciled by contacting the site and verifying the data.

Three types of analyses were conducted. A meta-analysis determined the average effect size of the intervention strategy. Individual time-series analyses provided supporting detail. Post-hoc analyses were used to evaluate the extent of selection bias.

The meta-analysis compared the percent reduction in injuries over baseline in each of the first, second, third, fourth, and fifth years after observations began. A paired t-test comparing baseline levels ($M=8.49$) to post-initiative levels after 1 year of observations ($M=6.24$) showed that, on the whole, the 73 sites achieved significant reductions in injuries ($t(72)=7.31$, $p<0.0001$). Table 3 shows that the average reductions from baseline were 26% in the first year, 42% from baseline in the second, 50% in the third, 60% in the fourth, and 69% in the fifth year after observations began. Note that the percentages are calculated sequentially over baseline. For example, the Year 2 percent reduction was calculated from the recordable rate in Year 2, not from the recordable rate for Years 1 and 2 combined.

Individual time-series analyses provided detail on each site's safety performance. Preliminary analyses were conducted to identify seasonality, trends, and other autocorrelations in the data. Trends were of particular concern because they frequently appear in time-series injury data. Therefore, all data, even if annual, were tested for trends using standard regression analysis.

Because of the large impact of identifiable trends on the subsequent time-series analysis, there was concern over the regression analyses' accuracy in detecting trends in the annual data. Although power is increased by the smaller variance in aggregated data, power is decreased by the availability of fewer degrees of freedom. To measure the impact of aggregating data on the ability to detect trends, each of the monthly data sets where trends had been detected in the baseline were aggregated into annual data sets and the trend test was repeated on the annual baseline data. Of the 14 monthly data sets with one or more years of baseline, trends were detected in the aggregated data in 11 cases (79%). This confirms that the annual data are less desirable than the monthly data, but that they usually do not mask trends. The quantity of information in the annual data was deemed useful, accurate, and sensitive enough to warrant inclusion in the study.

All data were tested for trends, and monthly data were tested for seasonality. In 56 cases, there was no evidence of serial dependence (trends or seasonality) in the data, which meant independent groups t-tests were appropriate. In 16 cases, a pre-existing trend required the use of a t-test of the residuals from the pre-initiative regression lines. The trends were detected in a regression analysis of the recordable rates over time using a two-tailed significance level of $p<0.05$. Seasonality was detected in one case using an analysis of autocorrelations with a two-tailed significance level of $p<0.05$. This site's

Table 3
Demographics, results for individual projects, and overall results

Site	Industry	DOB [a]	No. of employees	Variable tested	BL years	BL rate [b]	Sequential follow-up rate					Percent reduction from baseline					t (df)
							Y1	Y2	Y3	Y4	Y5	Y1	Y2	Y3	Y4	Y5	
1	Chem	01-86	500	RIR	2.00	6.00	5.64	4.68	1.22	1.87		6	22	80	69		2.40 (3)**
2	Chem	06-88	1000	RIR	2.42	5.88	3.06	2.44	1.57	1.21	1.11	48	59	73	79	81	4.43 (4)****
3	Chem	01-89	1100	RIR	1.00	7.77	4.04	1.68				48	78				3.26 (25)***
4	Chem	04-89	700	RIR	1.25	15.32	17.00	10.42	7.58	8.08		−11	32	51	47		3.05 (71)***
5	Chem	11-89	442	RIR	1.84	4.54	3.31	2.07	3.39	4.51	3.60	27	54	25	1	21	1.60 (68)*
6	Chem	10-90	500	TS	3.75	8.30	4.39	3.89	4.64	1.74	3.11	47	53	44	79	63	2.57 (81)***
7	Chem	11-90	1000	RIR	3.17	20.95	14.98	14.80				28	29				3.85 (54)****
8	Chem	09-91	400	RIR	4.00	12.65	8.54	5.77	3.16	2.06		32	54	75	84		3.33 (6)***
9	Chem	09-91	98	RIR	3.67	9.43	5.66	0.00	8.48			40	100	10			1.92 (5)*
10	Chem	11-91	186	RIR	3.08	3.18	1.27	2.26	2.36			60	29	26			1.34 (70)*
11	Chem	01-92	70	RIR	2.00	9.81	4.52	5.42	1.70	4.42	0.93	54	45	83	55	91	2.84 (27)***
12	Chem	09-92	485	RIR	2.67	1.60	0.39	0.58	1.34	0.57	0.43	76	64	16	64	73	2.64 (44)***
13	Chem	09-92	1125	RIR	4.00	3.52	4.28	1.86	2.27	1.72		−22	47	36	51		3.52 (94)***
14	Chem	01-93	1800	RIR	3.00	5.16	2.80	1.88	2.13			46	64	59			5.50 (5)****
15	Chem	06-94	540	RIR	3.00	6.72	6.83	3.62	3.25			−2	46	52			1.89 (74)**
16	Chem	11-94	700	REG	4.00	3.80	1.97	1.65				48	57				−1.19 (62)
17	Chem	02-95	625	RIR	4.00	7.22	7.28	4.73				−1	34				1.01 (21)*
18	Chem	03-95	51	RIR	3.94	3.60	2.04					43					0.75 (19)
19	Chem	05-95	550	REG	4.00	5.63	3.69	3.91				34	31				4.93 (72)****
20	Chem	11-95	93	RIR	4.00	6.10	6.45					−6					−0.43 (63)
21	Chem	04-96	450	RIR	4.00	2.48	2.42					2					0.08 (58)
22	Chem	05-96	335	RIR	4.00	4.28	4.17					3					0.73 (33)
23	Electr	10-91	400	RIR	2.75	11.28	4.50	3.00	3.30			60	73	71			5.84 (4)***
24	Electr	02-92	110	RIR	4.00	17.00	15.20	11.10	7.10			11	35	58			2.30 (4)**
25	Electr	05-94	250	REG	1.33	21.14	12.90	5.86	4.73			39	72	78			9.06 (52)****
26	Electr	05-96	1197	RIR	3.92	2.04	2.03					0					0.44 (59)
27	Food	07-92	200	RIR	1.25	10.88	3.46	5.56	3.91			68	49	64			2.37 (18)**

(Table continued on next page)

Table 3—*contd*

Site	Industry	DOB[a]	No. of employees	Variable tested	BL years	BL rate[b]	Sequential follow-up rate					Percent reduction from baseline					t (df)
							Y1	Y2	Y3	Y4	Y5	Y1	Y2	Y3	Y4	Y5	
28	Food	03-94	1452	RIR	4.00	15.08	17.42	14.81	10.80			−16	2	28			1.56 (86)*
29	Food	08-95	500	RIR	3.60	22.66	14.50	11.90				36	47				3.73 (4)**
30	Glass	11-93	406	REG	4.00	10.69	6.43	6.62	5.47			40	38	49			−0.45 (90)
31	Glass	04-94	400	RIR	3.25	9.21	4.13	2.46				55	73				4.61 (67)****
32	Glass	06-94	275	REG	4.00	3.34	5.07	3.78	1.54			−52	−13	54			−2.66 (83)
33	Glass	01-95	113	RIR	4.00	8.91	6.22	3.51				30	61				1.77 (72)**
34	Lumber	04-94	386	REG	3.92	16.89	11.62	6.21	4.76			31	63	72			−0.43 (75)
35	Lumber	06-94	615	RIR	1.41	10.07	8.93	12.12				11	−20				−0.03 (34)
36	Measur	06-96	420	RIR	1.42	6.38	4.45					30					0.49 (46)
37	Metal	06-90	2000	RIR	4.00	15.72	13.70	12.90				13	18				2.41 (5)**
38	Metal	11-90	350	RIR	1.83	18.29	12.19	11.97	13.57	10.44		33	35	26	43		1.03 (1)
39	Metal	04-91	3000	RIR	4.00	6.20	3.50	2.60				44	58				3.19 (4)**
40	Metal	11-91	70	RIR	2.83	15.41	8.59	3.74	7.43	2.63		44	76	52	83		4.79 (79)****
41	Metal	02-93	700	DIR	4.00	1.04	0.45	0.47	0.10	0.38	0.17	57	55	90	64	84	2.46 (8)**
42	Metal	11-95	620	RIR	4.00	4.02	7.34	8.20				−83	104				−3.37 (66)
43	Paper	09-91	190	RIR	1.67	8.79	3.05	7.65	3.92			65	13	55			2.13 (3)*
44	Paper	10-92	460	RIR	3.37	20.26	17.86	8.24				12	59				4.49 (42)***
45	Paper	11-92	255	RIR	3.92	10.09	7.68	4.93				24	51				2.25 (71)***
46	Paper	04-93	300	RIR	2.25	7.84	8.72	2.75	3.78			−11	65	52			1.78 (46)**
47	Paper	10-93	125	RIR	2.75	6.30	4.40	3.10				30	51				4.08 (3)**
48	Paper	01-94	360	RIR	4.00	5.03	3.90	1.24				22	75				2.51 (4)**
49	Paper	04-94	1100	REG	4.00	6.45	5.10	3.23				21	50				−1.46 (4)
50	Paper	07-94	412	REG	3.50	5.89	5.67	5.49				4	7				−2.13 (64)
51	Paper	10-94	650	RIR	3.92	12.62	7.50					41					1.75 (63)**
52	Paper	01-95	190	REG	4.00	14.09	10.63					25					−1.18 (64)
53	Paper	02-95	680	REG	4.00	11.86	7.38					38					3.64 (62)****
54	Paper	02-95	780	REG	4.00	14.00	7.82	12.02				44	14				5.36 (72)****
55	Paper	02-95	400	REG	4.00	10.48	7.05					33					0.31 (10)

(*Table continued on next page*)

Table 3—*contd*

Site	Industry	DOB[a]	No. of employees	Variable tested	BL years	BL rate[b]	Sequential follow-up rate					Percent reduction from baseline					
							Y1	Y2	Y3	Y4	Y5	Y1	Y2	Y3	Y4	Y5	t (df)
56	Paper	04-95	128	RIR	2.25	6.86	4.30	2.73				37	60				2.43 (42)***
57	Paper	12-95	324	RIR	3.92	8.39	4.37					48					1.91 (61)**
58	Petrol	11-90	140	RIR	3.84	5.37	3.13	3.45	2.43	0.91		42	36	55	83		3.90 (5)***
59	Petrol	09-91	360	RIR	2.42	4.44	2.74	3.82	2.70	2.83	1.39	38	14	39	36	69	2.00 (75)**
60	Petrol	03-93	900	REG	3.16	4.31	3.70	3.65	3.68	2.28		14	15	15	47		−5.85 (76)
61	Petrol	09-93	1000	REG	4.00	3.08	2.17	1.25	0.33			30	59	89			−1.39 (77)
62	Petrol	09-93	1400	RIR	3.67	2.91	2.13	1.84	1.88			27	37	35			2.88 (65)***
63	Petrol	05-94	1100	REG	4.00	4.90	2.60	1.99	3.07			47	59	37			9.34 (84)****
64	Petrol	05-94	192	RIR	1.58	3.55	3.38	3.14	4.16			5	12	−17			0.39 (55)
65	Petrol	08-94	250	RIR	3.97	10.96	4.97	5.07				55	54				4.39 (64)****
66	Petrol	11-94	245	RIR	1.85	2.91	2.40	2.19				18	25				0.38 (49)
67	Plastic	04-90	175	RIR	2.25	3.01	1.19	1.69				60	44				4.06 (4)***
68	Plastic	02-95	230	RIR	3.09	2.02	2.12	0.84				−5	58				0.39 (17)
69	Plastic	03-96	553	RIR	4.00	2.20	1.57					29					0.74 (59)
70	Service	02-95	1500	REG	4.00	24.18	29.08	18.21				−20	25				−2.38 (72)
71	Transp	10-89	187	REG	2.75	5.36	3.84	2.06	1.53	1.56		28	62	71	71		3.97 (5)***
72	Transp	01-94	109	RIR	4.00	9.14	6.22	4.90	4.71			32	46	48			2.29 (79)***
73	Transp	07-94	250	RIR	2.50	2.18	1.28					41					1.87 (19)**
Overall						8.49						26	42	50	60	69	7.31 (72)***

BL, baseline; Y, year; df, degrees of freedom; Chem, chemical; Electr, electrical; Measur, measuring instruments; Transp, transport; RIR, recordable rate; TS, time-series residuals; REG, regression residuals; DIR, disabling injury rate. *One-tailed $p<0.10$; **one-tailed $p<0.05$; ***one-tailed $p<0.01$; ****one-tailed $p<0.001$.

[a] Defined as the month in which behavioral observations began.

[b] Defined as the average US Occupational Safety and Health Administration recordable rate in the pre-intervention period (rate = cases × 200,000/h).

follow-up data were compared to baseline using an ARIMA (autoregressive-integrated moving average) time-series analysis (Box and Jenkins, 1976).

These site-specific results are reported in Table 3. To simplify reading of the table, sites are grouped by industry and listed in order of start date within each industry. Note that there is no gap between the baseline and follow-up periods. Also shown are the number of employees participating in the safety process, the variables tested, the baseline value of the dependent variable, the number of years available for the baseline, the percentage improvement in each follow-up year (sequentially over baseline), and the results of one-tailed individual time-series analyses.

The reader will note that the degrees of freedom on individual analyses vary from 1 to 94. The actual number of degrees of freedom for an individual test depends on the type of data (monthly, quarterly, or annual), the length of time available in the baseline and follow-up periods, and the specific test used (e.g. equal variance *t*-test, unequal variance *t*-test, time-series analysis). The 54 monthly and quarterly data sets have 10 or more degrees of freedom while the 19 annual data sets have eight or less.

Since it was discovered that the sample population differed from the target population on two demographic variables, size and longevity, post-hoc correlational analyses were conducted to examine the relationship between these variables and the percent reduction in injury in the first year. Although there was no relationship between number of employees and first year injury reductions, it was discovered that sites with more years of observations had greater reductions in injuries in the first year of their process than those with fewer years of observations ($r = 0.24$, $p < 0.01$).

One final post-hoc analysis was conducted to compare the effect size estimates in Year 1 between sites who shared data through the benchmarking service and sites who shared their data for other reasons. The 33 participants who received benchmarking services in exchange for their data experienced a 16% reduction from baseline in their first year. In contrast, the remaining non-benchmarking participants ($n = 40$) experienced a 35% reduction in their first year. This difference was statistically significant ($t(51) = -2.86$, $p < 0.01$).

6. Discussion

Since the 1930s safety managers have known that most injuries have a behavioral component. More recently, research has shown that effectively designed behavioral safety initiatives reduce incidents. This study quantified the effect size of a particular approach to behavior-based safety consisting of organizational assessment, behavioral observation, performance feedback, data analysis, action planning, and employee involvement. Results reported by the 73 facilities who implemented this behavioral approach to safety demonstrated improvement ranging on average from 26% in the first year to 69% by the fifth year; measured as each year's performance compared to baseline.

The validity of these results depends upon the absence of (1) substantive differences between the 73 sites studied and the population of interest (external validity), and (2) viable alternative explanations for the improvements in safety performance across sites (internal validity).

The external validity of these findings would be compromised if substantive differences exist between the 73 participants who sent complete data and the remaining 156 who either sent something ($n=47$) or who did not send data ($n=109$). The greatest concern is whether or not only the successful sites sent data. If so, the results presented here are representative of only the successful sites, not of the target population.

Data collection methods bear on this issue. Sites participating in the benchmarking service informally report that they started benchmarking because they were struggling with their behavior-based safety processes and wanted to find ways to improve it. The benchmarking service, initiated in 1996, provides a useful mechanism to compare the results of sites that participate in the benchmarking service to those sites that shared data on first request or in exchange for the custom data analysis. One would expect sites struggling with their safety process to be motivated by benchmarking because it provides a mechanism to improve their processes. Conversely, sites that shared their data on first request or in exchange for custom analysis would be more motivated to do so by their own positive results. The participants who received benchmarking in exchange for their data experienced a 16% reduction from baseline in their first year, while the remaining non-benchmarking participants experienced a 35% reduction in their first year. This statistically significant difference confirms that data collection methods do have an effect on the results. It also confirms that both successful sites and struggling sites were represented in the study: over half of the participants in the present study came from the benchmarking center, and benchmarking participants were known to have participated in benchmarking because they were struggling with their processes and wanted help.

The effects of selection bias was also examined by comparing the sample to the population on number of employees, longevity, industry, union status, and perceptions of success of the behavior-based safety process. As noted, the 73-site sample did not differ significantly from the 156 non-participants in the target population on union representation, industry, and responses to the question "Do you view your process as a success?"

The sample had a higher average number of employees than the remainder of the target population. However, the post-hoc correlational analysis showed that effect size does not depend on the number of employees. While not conclusive because the post-hoc analysis was limited to the sample, this suggests that the difference in size of implementation would not have affected the effect size estimates reported here.

The sample also had significantly lower longevity than the remainder of the population. Post-hoc correlational analysis showed that sites with more years of observations had greater reductions in injuries in the first year. Because sites with more years of observations were underrepresented in the sample, we expect that the effect sizes reported in the present study would underestimate the effect size in the population.

External validity of these findings is clearly limited by the inclusion criterion that each steering committee must have been trained directly by a consultant from the authors' firm. This criterion has two major implications. First, it excludes an important strategy for implementation that utilizes internal consultants. Further research is needed to evaluate the effect size of the behavior-based safety process using this model. Second, the present research was limited to organizations that used the consulting services of a single firm. The results of organizational change efforts are significantly influenced by the way in which they are implemented (Katz, 1993; Spector and Beer, 1994; Hidley, 1999). Because different consulting companies employ different implementation strategies, we do not necessarily expect the results of the present study to generalize to all behavior-based safety implementations.

External validity may also be limited by the other two inclusion criteria: each site was located in North America and had completed 1 year of behavioral safety observations. It is unknown to what extent these results will generalize to other groups; this is an important area for future research.

As is often the case in applied research, true experiments are not feasible, which means there may be explanations for changes in the dependent variable other than the independent variable. In the present study, one could ask whether or not the commitment to implement a behavior-based safety process is not enough to produce the observed improvements in the recordable injury rate. Whenever the independent variable is an entire package of interventions, it is difficult, if not impossible, to separate the contributions of each piece. Short of a true experiment, time-series designs such as the multiple-baseline can shed light on the issue by providing information on the timing of the change in safety performance. Krause (1999) studied the effects of the same behavior-based safety process evaluated in the present research on five organizations using a multiple-baseline design. He found that the observed improvement in incident rates coincided with the start of observations not with the commitment of resources, which typically occurred 6 months earlier. In addition, there are numerous studies demonstrating that changes in safety behavior coincide directly with behavioral intervention efforts (e.g. Komaki et al., 1978; Sulzer-Azaroff and de Santamaria, 1980; Sulzer-Azaroff et al., 1990).

Another potential threat to internal validity would be a global trend toward improved safety performance. According to the US Bureau of Labor Statistics, the average occupational injury and illness rates across private industry have neither increased nor decreased consistently in the last 10 years (US Department of Labor, 1997). Manufacturing industries (a description which covers most of the sites for the present study) show a slight downward trend. The average occupational injury and illness rates for this group decreased from 13.1 in 1989 to 10.6 in 1996. Fig. 1 compares this national trend to the baseline and follow-up periods of the 73 sites studied. It shows that the baseline average across all 73 sites was not decreasing with the national trend. This means the 73 sites would have been expected to continue at the same level of performance without intervention, while the national average of manufacturing organizations was expected to continue to improve. The follow-up period for the 73 sites showed a steeper downward trend than the national average of manufacturing companies. Fig. 1 provides some evidence that the 73 sites' improvement is not likely to be solely a result of the global improvement in recordable

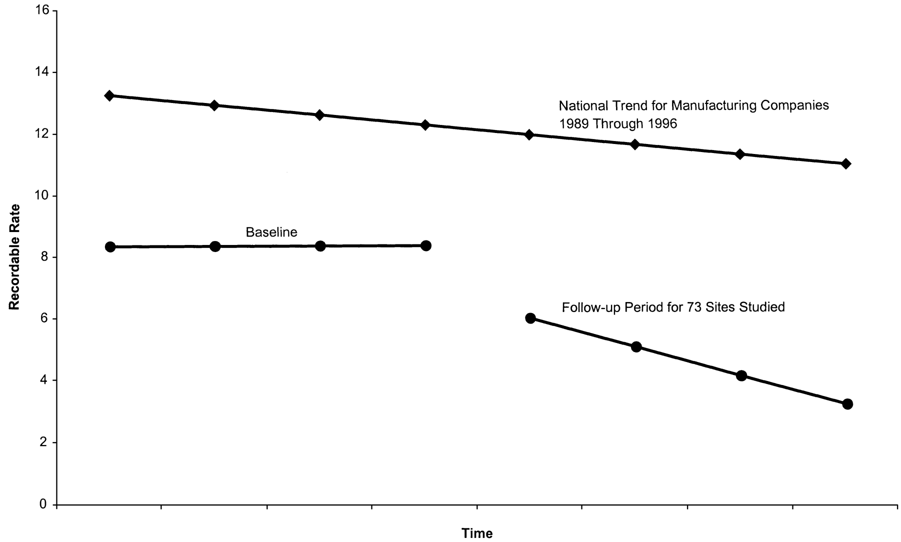

Fig. 1. Comparison of national trend in manufacturing companies to baseline and follow-up periods of 73 sites studied.

rates. Unique strengths of this study come from the large number and variety of companies who participated. Statistically significant reductions were demonstrated at different times, in different settings, for different industries, and over an 11-year span. The patterns of injuries were assessed over long periods of time before and after implementation, showing that in some cases the reductions in injuries not only endure but also improve over time.

The present research is part of an ongoing effort to evaluate the results of behavior-based safety initiatives using the approach described here. It is unique in that it was a safety consulting organization's effort to systematically evaluate the results of its clients. It demonstrates that the difficulties with data collection in the field can be overcome, and the benefit of the research prevailed over the many other pressures that discourage practitioners from conducting outcome research. Future research is needed to provide estimates of the effect size with larger and more representative samples, as are controlled experiments and component analyses.

Acknowledgements

We are grateful to Dr Beth Sulzer-Azaroff, Dr John Austin, and Dr A.R. Hale for their helpful suggestions in preparing this manuscript.

References

Box, G.E.P., Jenkins, G.M., 1976. Time-Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Chhokar, J.S., Wallin, J.A., 1984. Improving safety through applied behavior analysis. Journal of Safety Research 15(4), 141–151.

Deming, W.E. 1986. Out of the Crisis. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, MA.

Earley, C., 1988. Computer-generated performance feedback in the magazine-subscription industry. Organizational Behavior and Human Decision Processes 41, 50–64.

Guastello, S.J., 1993. Do we really know how well our occupational accident prevention programs work? Safety Science 16, 445–463.

Hidley, J.H. 1998. Seven critical success factors for behavior-based safety: avoiding common pitfalls and achieving real gains. Professional Safety 43(7), 30–34.

Katz, A.S., 1993. Eight TQM pitfalls. Journal for Quality and Participation 16(4), 24–27.

Killimett, P., 1991a. The identification of critical behaviors: the first step in a behavior-based safety process. Tappi Journal 74(10), 251–253.

Killimett, P., 1991b. The behavior-based approach to safety: a clear picture, an effective plan. Tappi Journal 74(8), 141–143.

Killimett, P., 1992. Behavioral observation: gathering data for continuous safety improvement. Tappi Journal 75(2), 310–312.

Komaki, J., Barwick, D., Scott, L.R., 1978. A behavioral approach to occupational safety: pinpointing and reinforcing safety performance in a food manufacturing plant. Journal of Applied Psychology 63(4), 434–445.

Komaki, J., Heinzmann, A.T., Lawson, L., 1980. Effect of training and feedback: component analysis of a behavioral safety program. Journal of Applied Psychology 65(3), 261–270.

Kopelman, R.E. 1986. Managing Productivity in Organizations: A Practical, People-Oriented Perspective. McGraw-Hill, New York.

Krause, T.R. 1989. Critical Mass Approach to Accident Prevention. Behavioral Science Technology, Inc. Ojai, CA.

Krause, T.R., 1992. Assessment: tailoring the behaviour-based safety process. The Safety and Health Practitioner 10(4), 12–14.

Krause, T.R. 1995. Employee-Driven Systems for Safe Behavior. Van Nostrand Reinhold, New York.

Krause, T.R. 1997. The Behavior-Based Safety Process: Managing Involvement for an Injury-Free Culture, 2nd Edition. Van Nostrand Reinhold, New York.

Krause, T.R. 1999. Skinner meets Deming: employee-driven behavioral methods (submitted).

Krause, T.R., Hidley, J.H., 1989. Behaviorally based safety management: parallels with the quality improvement process. Professional Safety 34(10), 20–25.

Krause, T.R., Hidley, J.H., Hodson, S.J. 1990. The Behavior-Based Safety Process: Managing Involvement for an Injury-Free Culture. Van Nostrand Reinhold, New York.

Krause, T.R., Hidley, J.H., Lareau, W., 1984. Behavioral science applied to industrial accident prevention. Professional Safety 29(7), 21–27.

Locke, E.A., Shaw, K.N., Saari, L.M., Latham, G.P., 1981. Goal setting and task performance: 1969–1980. Psychological Bulletin 90(1), 125–152.

McAfee, R.B., Winn, A.R., 1989. The use of incentives/feedback to enhance work place safety: a critique of the literature. Journal of Safety Research 20, 7–19.

Reber, R.A., Wallin, J.A., Chhokar, J.S., 1984. Reducing industrial accidents: a behavioral experiment. Industrial Relations 23(1), 119–125.

Salazar, N., 1989. Applying the Deming philosophy to the safety system. Professional Safety 34(12), 22–27.

Spector, B., Beer, M., 1994. Beyond TQM programmes. Journal of Organizational Change Management 7(2), 63–70.

Sulzer-Azaroff, B., de Santamaria, M.C., 1980. Industrial safety hazard reduction through performance feedback. Journal of Applied Behavior Analysis 13(2), 287–295.

Sulzer-Azaroff, B., Loafman, B., Merante, R.J., Hlavacek, A.C., 1990. Improving safety in a large industrial plant: a systematic replication. Journal of Organizational Behavior Management 11(1), 99–120.

US Department of Labor, 1986. A Brief Guide to Recordkeeping Requirements for Occupational Injuries and Illnesses (OMB No. 1220-0029). US Department of Labor, Washington, DC.

US Department of Labor, 1997. Bureau of Labor Statistics Data (on-line): http://stats.bls.gov/sahome.html.