

Evaluación de Programas de Intervención para la Prevención de Accidentes en Ambientes Laborales: Consideraciones Estadísticas

EVALUATION OF INJURY PREVENTION INTERVENTION PROGRAMS IN OCCUPATIONAL SETTINGS: STATISTICAL CONSIDERATIONS

Shrikant I. Bangdiwala

PhD Department of Biostatistics, and Injury Prevention Research Center University of North Carolina at Chapel Hill, U.S.A.

RESUMEN

Los programas de intervención para la promoción de la seguridad y la prevención de accidentes se utilizan comúnmente en ambientes ocupacionales. Por ejemplo, la Asociación Chilena de Seguridad (ACHS) periódicamente desarrolla e implementa campañas de seguridad en ambientes ocupacionales seleccionados. Los ambientes laborales pueden ser fábricas, oficinas o cualquier lugar en que haya trabajadores. Son variables en sus características y circunstancias, y así cada programa varía en cuanto a sus componentes, impulsados por necesidades locales, poblaciones de trabajadores y financiamiento. Es imperativo que tales esfuerzos sean evaluados. La evaluación puede ser realizada en una variedad de formas, desde resultados casuales, cualitativos y anecdóticos hasta evaluación científicamente objetiva y rigurosa empleando información cualitativa y cuantitativa. Este documento explora varios procedimientos de evaluación disponibles. Nuestro punto clave es que se necesita una variedad de enfoques, dado que distintos enfoques interesan a diferentes audiencias y propósitos.

(Bangdiwala S.2006. Evaluación de Programas de Intervención para la Prevención de Accidentes en Ambientes Laborales: Consideraciones Estadísticas. Cienc Trab Ene.-Marz.;8(19):31-36.

Descriptores: PREVENCIÓN DE ACCIDENTES, SEGURIDAD, SALUD OCUPACIONAL/ESTADÍSTICA Y DATOS, PROMOCIÓN DE LA SALUD, CHILE

ABSTRACT

Injury prevention and safety promotion intervention programs are commonly utilized in occupational settings. For example, the Asociación Chilena de Seguridad (ACHS) periodically develops and implements safety campaigns in selected occupational settings. Occupational settings may be factories, offices, or anywhere that workers are present. They are variable in their characteristics and circumstances, and thus each program varies as to its components, driven by local necessities, worker populations and funding. It is imperative that such efforts be evaluated. Evaluation can be performed in a variety of ways, from casual, qualitative and anecdotic results, to scientifically objective and rigorous evaluation using qualitative and quantitative information. This paper explores several evaluation procedures available. Our key point is that a variety of approaches is necessary, since different approaches appeal to different audiences and purposes.

Descriptors: ACCIDENT PREVENTION, SAFETY, OCCUPATIONAL HEALTH/STATISTICS AND NUMERICAL DATA, HEALTH PROMOTION, CHILE

INTRODUCCIÓN

Los programas para reducir accidentes y promover la seguridad en el trabajo son, por definición, implementados en fábricas o ambientes ocupacionales y afectan a todos los trabajadores. Las instituciones pueden elegir por sí mismas implementar una determinada campaña o programa de seguridad, y sus motivaciones varían. Las empresas son bastante variables en su tamaño, lugar, población, tasas de accidentes, circunstancias y características. Es

por lo tanto lógico y esperado que los programas de prevención de lesiones y de promoción de la seguridad que ellos implementen sean bastante variados en cuanto a sus componentes. Además, son en gran medida impulsados por necesidades locales, población de trabajadores atendida y financiamiento.

Los programas de prevención de accidentes y promoción de la seguridad son grandes esfuerzos, involucran un gran segmento de la población, la movilización de recursos y, por lo tanto, requieren que se integre un componente de evaluación en cada programa individual. Se pueden considerar numerosas posibilidades, desde estudios cualitativos, evaluaciones de procesos (Mitchell 1998), hasta estudios cuantitativos tales como diseños cuasi-experimentales y ensayos de intervención en la comunidad (Menckel 1999).

La evaluación puede realizarse en una diversidad de formas, desde resultados casuales, cualitativos y anecdóticos, hasta una evaluación científicamente objetiva y rigurosa. Este manuscrito explorará algunos de los procedimientos de evaluación disponibles y sus consideraciones estadísticas. Nuestro punto clave es que se necesita una variedad de enfoques, dado que diferentes perspectivas interesan a diferentes audiencias y objetivos.

Correspondencia / Correspondance:

Shrikant I. Bangdiwala

Department of Biostatistics, and Injury Prevention Research Center

University of North Carolina at Chapel Hill, U.S.A

e-mail: kant@unc.edu

Recibido: 29 de enero de 2006 / Aceptado: 21 de febrero de 2006

OBJETIVOS PARA LAS EVALUACIONES

En el campo de la salud pública, la evaluación apunta a responder preguntas como:

- ¿Ha llevado el programa de intervención en seguridad ocupacional a reducciones en la cantidad de los accidentes, la gravedad de éstos y sus costos?
- ¿Ha llevado el programa de intervención en seguridad ocupacional a un mayor conocimiento, un cambio de actitudes y de conductas relacionadas con riesgos de accidentes?
- ¿Ha tenido el programa de intervención en seguridad ocupacional un impacto para los trabajadores, las empresas, las comunidades, las regiones, y naciones?

Preguntas como las anteriores se refieren a la efectividad de un programa en términos de resultados medibles o su impacto. Este enfoque es lo que a menudo hacen los evaluadores externos al programa. La investigación de evaluación es así a menudo llamada investigación de resultados, dado que uno se preocupa de evaluar el impacto o resultados después que se ha llevado a cabo una intervención o programa específico. Dicha evaluación hecha por pares y gente externa es para justificar que lo que uno está haciendo merece la inversión de tiempo, de dinero y otros recursos. Otras preguntas, tales como, ¿qué puede hacerse para mejorar la efectividad de la intervención? conducen a la evaluación concomitante del proceso de implementación de la intervención mientras éste se está llevando a cabo, y tienen el propósito de garantizarle a uno que está en el camino correcto. Esto conduce a los conceptos de “gestión total de la calidad” y al “auto-mejoramiento continuo”. Los “resultados” para estos tipos de preguntas se relacionan a menudo con los procesos de la intervención y no con los llamados resultados definitivos o “duros”. Ejemplos incluyen la aceptación de la intervención por los grupos claves, la facilidad o dificultad en llevar a cabo los cambios ambientales, la participación de trabajadores en actividades del programa, etc. Un buen conjunto de ejemplos de evaluaciones de procesos es presentado por Steckler y Linnan (2002). Los programas deberían considerar ambos tipos de enfoques, el continuo llamado “evaluación de procesos” así como la “evaluación de resultados”.

Un elemento a menudo no considerado en la investigación de evaluación es la valoración del programa por parte de los que participaron en el estudio de intervención, es decir, los trabajadores —uno de los principales grupos de interés. En los estudios de seguridad ocupacional, es necesario ofrecer una tranquilidad continua a los líderes laborales y a los trabajadores de que su tiempo y esfuerzos están llevando a resultados significativos tanto a nivel individual como de la empresa. Esto podría llamarse “evaluación social”.

Entonces, podemos decir que hay tres impactos que una intervención puede tener y deben ser por lo tanto objetivos de un esfuerzo de evaluación: impacto en los indicadores de accidentes, en el proceso y sobre aspectos sociales. Evaluar éstos requiere diferentes enfoques.

ALTERNATIVAS DE INVESTIGACIÓN DE EVALUACIÓN

La investigación de evaluación es la “aplicación sistemática de procedimientos de investigación social para evaluar la conceptualización, diseño, implementación y utilidad de los programas de intervención social” (Rossi y Freeman 1993). Los procedimientos

pueden clasificarse bajo diferentes nomenclaturas. (Menckel 1999) los desglosa en tres tipos: aditivo, formativo y diagnóstico. La evaluación aditiva se lleva a cabo retrospectivamente después que se ha completado una intervención, mientras que la formativa es contemporánea al programa de intervención. La evaluación por adelantado es lo que ella llama “diagnóstico” y apunta a la planificación sistemática. La evaluación formativa es para apoyar y guiar la intervención, mientras que la evaluación aditiva es para controlar y evaluar el impacto de la intervención.

Los datos recopilados para propósitos de evaluación pueden ser cualitativos o cuantitativos. La información cualitativa implica obtener datos detallados, en profundidad, procedentes de unos pocos informantes claves, y frecuentemente no puede ser generalizada a una población objetivo. Proporciona información valiosa acerca del lenguaje, contexto y relaciones entre ideas. La información cuantitativa proviene de la recopilación formal o estructurada de datos de un número relativamente grande de grupos de interés y si es recopilada bajo muestras de probabilidad, proporciona resultados que son generalizables a la población objetivo. Ambos tipos de información tienen fortalezas y debilidades. Utilizan diferentes diseños de estudio, desde discusiones de focus group, entrevistas en profundidad, encuestas de preguntas abiertas y diarios para información cualitativa, hasta encuestas muestrales, revisión de registros y diseños de investigación observacionales y experimentales tales como estudios transversales, estudios de cohorte y estudios de intervención para información cuantitativa.

Dado que los objetivos de la evaluación pueden variar desde valorar los resultados, el proceso y/o el impacto social, los métodos para evaluación también varían. Si se está interesado en los resultados, uno debe establecer ya sea recopilación continua de datos acerca de la mortalidad asociada a accidentes y morbilidad tales como métodos de vigilancia, o realizar evaluaciones periódicas en la forma de encuestas o estudios transversales. Los resultados del proceso se refieren a elementos de la realización de un estudio, tales como el desarrollo de la voluntad política por parte de los líderes laborales y de la empresa, colaboración intersectorial, apoyo financiero y político del programa y el interés y participación de los trabajadores en diversos aspectos del programa. El último puede ser medido con evidencias externas cuantificables de cambios tales como un mayor uso de casco en el lugar de trabajo, cambios en las prácticas y conductas de seguridad, etc. Otros resultados del proceso se prestan para entrevistas en profundidad a una cantidad limitada de individuos, y así generan información cualitativa.

El impacto social de una intervención es más difícil de definir, por lo tanto, de evaluar, y se presta principalmente para enfoques de información cualitativa. Por ejemplo, medir la “imagen” que un programa tiene en la sociedad es un aspecto importante de evaluación, pero difícil de cuantificar.

GRUPOS CLAVES DE INTERÉS

Es importante identificar por qué se está realizando la evaluación y cómo se han de utilizar sus resultados. El objetivo usual tras una evaluación es medir el impacto de la intervención sobre la reducción de la cantidad y/o gravedad de los accidentes, pero la forma en que los resultados de la evaluación serán empleados depende de los grupos de interés claves. En cualquier proceso de

evaluación, es esencial identificar estos grupos. Los grupos de interés son individuos o instituciones que responden a las siguientes dos preguntas: ¿para quién se ha de realizar la evaluación?, y ¿quién ha de realizar la evaluación? Los grupos claves de interés identificados por (Menckel 1999) incluyen:

- Políticos, comisionados del proyecto, directores de la municipalidad
- Juntas, administraciones y oficinas municipales
- Organizaciones voluntarias
- Dirección de programa y grupos asesores
- Personas o instituciones que trabajan en el programa
- Residentes del municipio o comunidad
- Investigadores
- Otras personas o instituciones interesadas tales como otros municipios y comunidades
- Personas o instituciones internacionales interesadas

Los grupos de interés pueden ser internos o externos al programa, como las agencias reguladoras, la administración de la empresa, los trabajadores, los pares científicos o la comunidad en general. Los grupos de interés internos están interesados en los resultados de tipo proceso y el impacto social de la intervención, mientras que los grupos de interés externos ya sea controlan el financiamiento o la opinión de los pares sobre la eficacia del programa y a menudo “exigen” los llamados resultados más duros de impacto cuantitativo tal como una reducción en las tasa de lesiones.

IMPACTO CUALITATIVO Y/O RESULTADOS

El resumen y la presentación de información cualitativa puede ser un sólido complemento de los resultados cuantitativos típicos. Ellos pueden poner en contexto la información cuantitativa, ya que ayudan a explicar e interpretar los resultados. Eventos anecdóticos o entrevistas en profundidad de los grupos de interés clave son elementos importantes de cualquier evaluación, y son especialmente importantes en los informes de avance para una actividad de intervención de seguridad ocupacional en curso. Los informes de grupos claves de interés pueden proporcionar una motivación continua para la actividad, mejorar la aprobación y ayudar a asegurar el financiamiento permanente. Los resultados del proceso pueden ayudar en el mejoramiento continuo de la actividad.

La información cualitativa por sí misma no es considerada suficientemente sólida en muchas circunstancias, y debería ser complementada mediante simples medidas resumen de tendencias globales en algunos de los resultados tradicionales tales como tasas de mortalidad por accidentes y de morbilidad. Los políticos también necesitan números para “justificar” su apoyo continuo a una actividad.

RESULTADOS CUANTITATIVOS

Análisis de tendencia en una serie cronológica

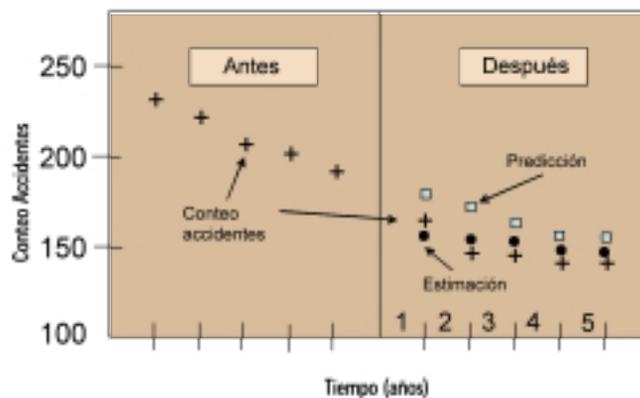
Cada programa de seguridad individual puede examinar indicadores claves —tasas de morbilidad y mortalidad por accidentes, distribución accidentes según tipo y gravedad, costos directos e indirectos— en el tiempo. Desde un punto de vista estadístico, sería mejor tener mediciones muy frecuentes, pero dado que por

naturaleza los accidentes son “eventos raros”, la frecuencia de las mediciones es probablemente mejor observada en un período de 6 meses o un año. Este enfoque proporciona una evaluación visual de las tendencias cronológicas, pero requiere un largo período de tiempo a fin de evaluarlos, y necesita empresas de comparación a fin de evaluar si las tendencias observadas en las empresas con programas de intervención son diferentes de las tendencias de la industria.

Estudios antes-después

Esta metodología se usa bastante, dado que la misma empresa sirve como su propio control o comparación, reduciendo así la variabilidad inherente si uno comparara una empresa con una empresa de control diferente (Hauer 1997). En tales estudios, el análisis se hace básicamente comparando la tendencia previa en la empresa con la tendencia observada, en ella con posterioridad a la introducción de la intervención (Figura 1).

Figura 1. Serie Cronológica Tipo Antes-Después Hipotética. La línea vertical indica la introducción de la intervención



Observe que para este enfoque uno necesita varias mediciones antes de la introducción de la intervención, a fin de poder establecer la tendencia en curso con alguna precisión, así como varias mediciones después de la introducción de la intervención a fin de establecer una nueva tendencia para comparación con la anterior.

Empresas para comparación

La necesidad de contar con una empresa para comparación adecuada es esencial para una evaluación válida de la efectividad de una intervención. Así, un programa de seguridad ocupacional implementado en una determinada empresa debería compararse contra una empresa de control adecuada con respecto a las variables de resultado de interés. Puede que el problema no sea la dificultad de encontrar una empresa comparable para que sirva como control, sino el hecho de que este enfoque está repleto con la posibilidad de sesgo de selección, en que la empresa que elige participar en el programa de seguridad ocupacional tiene características específicas que pueden afectar sus resultados —p. ej., la disposición para participar implica voluntad política de parte de los trabajadores y administración de la empresa y así los resultados pueden deberse más a esta buena disposición que a los efectos de la intervención per se. Lo que se necesita es un proceso

de asignación aleatoria a la intervención o control, un proceso que en la mayoría de los casos puede no ser práctico (Bangdiwala 2001).

Estudios de intervención minimizada o aleatoria

Si es posible asignar en forma aleatoria empresas a intervención o control, uno obtiene el rigor estadístico necesario para probar la hipótesis de la eficacia de la intervención. Esto es visto como el más sólido diseño de investigación metodológica para evaluación (Bangdiwala 2003). Sin embargo, encontramos muchos obstáculos en este enfoque. Primero, las empresas deben estar dispuestas a ser asignadas al brazo de control, y las empresas motivadas puede que no deseen tal asignación. Una alternativa para resolver esto es hacer que el brazo de control sea también el de "intervención retardada", de manera que todas las empresas efectivamente obtengan la intervención, sólo que algunas la obtienen más tarde que otras, permitiendo a los investigadores evaluar el efecto de la intervención de una manera controlada y aleatoria antes que las empresas de control obtengan la intervención. Esto es difícil, dado que las empresas ansiosas pueden no estar dispuestas a ser demoradas mucho tiempo y el efecto de la intervención puede no ser evidente hasta después de un periodo de tiempo prolongado.

Otro problema con los diseños de estudio de asignación aleatoria es que uno tiene relativamente pocas empresas en tales estudios, y a menos que se utilice un procedimiento de asignación aleatoria restringida, es probable que el grupo de intervención de empresas sea diferente del grupo de empresas control en factores que pueden estar relacionados con los resultados de interés. Uno puede utilizar procedimientos restrictivos tales como selección aleatoria pareada, pero puede ser difícil encontrar empresas adecuadas para aparear si hay muchos parámetros importantes que considerar. La minimización es otro enfoque de la asignación, la cual, aunque no se utiliza con frecuencia, ofrece la posibilidad de equilibrar los brazos en la línea de base y aun obtener el beneficio de la asignación aleatoria (Scott et al. 2002).

Modelos multiniveles para análisis

Recientemente, el campo de la estadística aplicada se ha preocupado del llamado modelo multinivel. La teoría para tales modelos proviene de los de curva de crecimiento de los años 1950 para datos longitudinales y los diseños ANOVA de parcelas divididas para diseños de datos agrupados. Esta metodología para generar modelos de regresión, llamada en algunos campos, modelaje jerárquico es adecuada, cuando la información se recopila en diversos niveles que están anidados o jerarquizados (Snijders y Bosker 1999). Por ejemplo, en estudios de intervención de seguridad ocupacional, uno recopila información al nivel de la empresa e información a nivel del trabajador. La información a nivel del trabajador se considera anidada dentro de la empresa, y la información a nivel de esta última es compartida por todos los trabajadores de una determinada empresa.

La naturaleza jerárquica del anidamiento lleva al asunto estadístico de los datos correlacionados, en el sentido de que la información a nivel de trabajador o individuo está correlacionada debido al hecho que los trabajadores comparten características del nivel de la empresa. Si uno fuera a analizar la información cuantitativa de variables de nivel individual o variables de ambos niveles de varias empresas, uno debe considerar la correlación intra-clase entre observaciones dentro de las empresas. Esta correlación afecta las estimaciones de varianza de las estima-

ciones de efecto de la intervención. Durante la etapa de planificación de tales estudios, uno debe decidir el número de empresas (k) y el número de trabajadores dentro de las empresas (m) a incluir en el estudio. El tamaño de la muestra total es $n = mk$. La estimación del tamaño necesario de la muestra debe ser inflada multiplicándola por el *efecto de diseño*, que es

$$1 + (m-1)\rho,$$

donde ρ es el Coeficiente de Correlación Intraclase o CCI, resultando en el valor inflado

$$n^* = n [1 + (m-1)\rho].$$

El CCI es realmente la razón entre el componente de varianza entre empresas y la varianza total. Note que si $m=1$ o $\rho=0$ entonces no hay efecto del diseño, y el tamaño de la muestra no es aumentado.

El efecto de diseño también debe ser considerado en la etapa de análisis del estudio, donde afecta la varianza de los estimadores de manera similar, llevando a ser llamado el Factor de la Inflación de la Varianza (FIV). Así, las varianzas de las estimaciones deben ser multiplicadas por el efecto diseño, llevando a una menor significancia dado que los tamaños de los efectos [estimaciones/ desviación estándar] resultarán aminoradas.

Además de la información multinivel debida al anidamiento jerárquico de los trabajadores dentro de las empresas, nosotros típicamente tenemos medida repetidas longitudinales en el tiempo, y así los datos son realmente correlacionados no solamente en espacio (debido a la agrupación de información de trabajadores dentro de las empresas) sino que también en el tiempo (debido a la repetición de medidas en el tiempo). Estos modelos tendrían el siguiente formato genérico:

$$Y_{itk} = \beta_{0k} + \sum_{j=1}^p \beta_{1kj} X_{itkj} + \beta_{2k} t + \sum_{j=1}^q \beta_{3kj} Z_{itkj} + \sum_{j=1}^r \beta_{4kj} G_{itkj} + \epsilon_{itk}$$

donde el resultado continuo Y se mide para el k-esimo trabajador de la i-esima empresa en el punto temporal t, y uno tiene variables independientes (o descriptores) a nivel del trabajador (X y Z) o a nivel de la empresa (G). Podemos considerar algunos de los efectos de nivel de empresa como efectos aleatorios (Z), y podemos incluir valores separados para cada empresa de estos efectos (intercepto o pendiente aleatorios). Observe que dada la naturaleza longitudinal de los datos podemos también incluir un efecto fijo para el tiempo.

Observe también que dado el hecho que los accidentes son (afortunadamente) eventos raros, la variable de respuesta Y es a menudo un resultado binario que no ocurre con frecuencia. El modelo de regresión toma la forma de regresión logística o modelos multinivel de regresión de Poisson (Snijders y Bosker 1999).

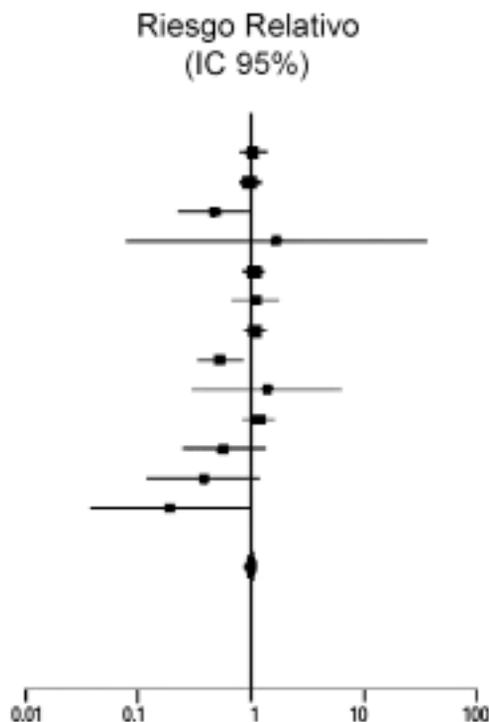
Meta-análisis de múltiples empresas

El uso de meta-análisis es una metodología común en investigación clínica para combinar los efectos de varios estudios similares. Uno puede agregar los resultados de los estudios ya sea combinando los niveles de significación o combinando los tamaños del efecto, tomando en

cuenta la “calidad” de los estudios. A menudo esto se hace usando una estimación de variabilidad (por ej. El error estándar de la estimación de la media), y se representa con lo que se llama “el diagrama de bosque” (Figura 2).

Figura 2.

Diagrama de Bosque Hipotético. Estimaciones de parámetro e intervalos de confianza de 95% para varios estudios, el tamaño de efecto combinado e intervalo confidencial de 95%.



El diagrama de bosque muestra la estimación y los intervalos confidenciales de 95% de confianza para el parámetro de interés (efecto de la intervención, tasa de accidentes, etc.) como barras horizontales, y un resumen o estimación de efecto agregada en la forma de un diamante junto con su intervalo de confianza de 95%. El largo de cada barra horizontal es una medida de la variabilidad inherente en el estudio en particular, a menudo indicio de tamaño de muestra pequeño y/o poca calidad. Observe que el intervalo de confianza de 95% para el efecto agregado del riesgo relativo es mucho más estrecho que los intervalos correspondientes a los estudios individuales.

También es posible agregar datos a través de las empresas para variables no continuas. Digamos que estamos interesados en demostrar que la intervención afecta un resultado medido a nivel binario en n sujetos, y supongamos que tenemos el valor antes y después en los mismos n individuos. La prueba apropiada para

una sola empresa es la prueba de McNemar, basado en la tabla de contingencia de 2x2 siguiente.

Tabla 1.

		Después		
		Malo	Bueno	
Antes	Malo	A	B	n_{MM}
	Bueno	C	D	n_{MB}
		n_{AD}	n_{AG}	n

La prueba compara los elementos fuera de la diagonal (B y C) en la tabla 2x2, buscando asimetría como un indicio de impacto. La estadística de la prueba tiene una distribución ji cuadrado de 1 grado de libertad. El método Mantel-Haenszel para combinar tablas 2x2 puede utilizarse para obtener una indicación global del impacto a través de varios programas de seguridad ocupacional.

Si el resultado es una variable continua o categórica medida solamente al nivel de la empresa, digamos la frecuencia de accidentes en un punto dado del tiempo después de la intervención, uno puede tratar las observaciones de los diferentes programas de seguridad ocupacional como estadísticamente independientes ya que ellos son un resumen estadístico para cada empresa. Uno puede entonces examinar la relación de factores de nivel empresa sobre estos resultados vía técnicas de regresión estándar, ya sea regresión lineal múltiple para respuestas continuas o modelos logarítmicos lineales (tales como regresión logística) para respuestas categóricas. El problema con este enfoque es que la cantidad de observaciones es relativamente poca y así tales modelos carecen de suficiente poder estadístico. Si modelamos respuestas a nivel del trabajador o individual, aumentamos la potencia, pero debemos utilizar los modelos mixtos o multinivel, como se describió anteriormente, para considerar adecuadamente la naturaleza correlacionada de los datos jerárquicos.

CONCLUSIONES

Todos los programas de intervención deberían ser evaluados. Es mejor una combinación de métodos cualitativos y cuantitativos a fin de satisfacer las diversas necesidades de evaluación y abordar los diferentes grupos de interés. En el escenario de la seguridad ocupacional, los programas de intervención diseñados para promover la seguridad y reducir los accidentes, normalmente se planifican en lo que puede llamarse un diseño de estudio cuasi-experimental, ya que las empresas son asignadas a brazos de intervención y luego se analizan los resultados de trabajadores anidados dentro de las empresas. El análisis cuantitativo que evalúa la efectividad de tales estudios debe considerar el efecto diseño. Los modelos de regresión para comprender la relación de muchas co-variables posibles deben ser modelos multinivel. Finalmente, si se agrega información cuantitativa de múltiples programas de intervención de seguridad ocupacional, se recomienda métodos meta-analíticos.

REFERENCIAS

- Bangdiwala SI. 2001. Statistical considerations for the design, conduct and analysis of the efficacy of safe community interventions, *Injury Control & Safety Promotion* 8:91-97.
- . 2003. Scientific considerations in injury research methodology, in *Second Advanced Nordic-Baltic Course on Safety Promotion and Injury Prevention Research in Toila, Estonia*, Ekman R, ed., Karolinska Institutet report 2003:2, 90-98.
- Hauer E. 1997. *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. New York: Pergamon Press, Elsevier Science, Inc.
- Menckel E. 1999. Evaluation research – Approaches, models and methods for evaluating safety promotion, In: *Safety Promotion Research*, Laflamme L, Svanstrom L, Schelp L, eds., Department of Public Health Sciences, Karolinska Institutet, Stockholm.
- Mitchell CS. 1998. Evaluating occupational health and safety programs in the public sector. *Am J Ind Med*. 34:600-6.
- Rossi PH, Freeman HE. 1993. *Evaluation: A Systematic Approach*. 5th ed Newsbury Park: SAGE Publications.
- Scott NW, McPherson GC, Ramsay CR, Campbell MK. 2002. The method of minimization for allocation to clinical trials: A review, *Controlled Clinical Trials* 23:662-74.
- Snijders TAB, Bosker R. 1999. *Multilevel Analysis : An Introduction to Basic and Advanced Multilevel Modeling*. London: SAGE,.
- Steckler A, Linnan L. 2002. *Process Evaluation for Public Health Interventions and Research*: San Francisco: Jossey-Bass, John Wiley.