

La evaluación del proceso en relación con la evaluación de la efectividad: experiencias de programas en el medio escolar

Carles Ariza^{a,b,c,*}, Joan R. Villalbí^{a,b,c,d}, Francesca Sánchez-Martínez^{a,b} y Manel Nebot^{a,b,c,d}

^aAgència de Salut Pública de Barcelona, España

^bCIBER de Epidemiología y Salud Pública (CIBERESP), España

^cInstituto de Investigaciones Biomédicas Sant Pau (IIB Sant Pau), Barcelona, España

^dDepartamento de Ciencias Experimentales y de la Salud, Universitat Pompeu Fabra, Barcelona, España

RESUMEN

Palabras clave:

Programas preventivos
Salud en la escuela
Evaluación de programas
Evaluación de proceso
Evaluación de la efectividad
Estudios cuasiexperimentales

La evaluación de las intervenciones de salud pública suele centrarse en la calidad del diseño y de los métodos de investigación, pero mucho menos en la calidad de la intervención o la evaluación del proceso. En la evaluación del proceso de las intervenciones preventivas escolares, un aspecto clave es el grado de cumplimiento, o dosis de la intervención efectivamente realizada, así como la fidelidad con respecto al protocolo. Además, es muy útil valorar las variables intermedias, como son las que influyen en la conducta que se quiere prevenir. En este trabajo se presentan los conceptos básicos sobre el tema, con ejemplos de evaluación de la efectividad de intervenciones preventivas realizadas en el medio escolar. Se describen principalmente estudios cuasiexperimentales de evaluación de programas impulsados por los servicios de salud pública de la ciudad de Barcelona. Se presenta información de la evaluación del proceso de programas de prevención en la escuela secundaria que han sido objeto de evaluación formal de su efectividad, aportando datos de la evaluación del proceso. Los ejemplos proceden de programas de prevención de la infección por el VIH y de las relaciones sexuales no protegidas (programa PRESSEC), y de prevención del consumo de sustancias adictivas (programas PASE, PASE.bcn y xkpts). Estos ejemplos muestran cómo el proceso influye en los resultados y el impacto logrados. La correcta planificación de la evaluación del proceso es fundamental para obtener indicadores válidos que permitan determinar, en la evaluación de la efectividad de la intervención, las estrategias más eficaces para obtener unos resultados positivos.

© 2011 SESPAS. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Process evaluation in relation to effectiveness assessment: experiences with school-based programs

ABSTRACT

Keywords:

Preventive programmes
Health school
Programme evaluation
Process evaluation
Effectiveness evaluation
Quasi-experimental studies

Evaluation of public health interventions usually focusses on the quality of design and research methods, and less on the quality of the intervention or process evaluation. In process evaluation of school-based interventions, key issues are how completely the intervention is carried out and adherence to the protocol. In addition, exploration of intermediate variables, such as those that influence (and often predict) preventable behavior, is highly useful. This article describes the basic concepts in this topic, using examples of the effectiveness of some preventive interventions carried out in schools. The interventions discussed were mainly quasi-experimental studies, based on data from programs promoted by public health teams in the city of Barcelona. Data from process evaluation of preventive programs in secondary schools that underwent formal assessment of their effectiveness is provided. The examples are drawn from preventive programs of HIV infection or unprotected sexual intercourse (PRESSEC program) and drug consumption prevention (the PASE, PASE.bcn and xkpts programs). These examples show why the intervention process influences the impact of the programs and their results. Thorough planning of process evaluation is essential to obtain valid indicators that will identify, in the effectiveness evaluation of the intervention, the most efficacious strategies to obtain positive outcomes.

© 2011 SESPAS. Published by Elsevier España, S.L. All rights reserved.

*Autor para correspondencia.

Correo electrónico: cariza@aspb.cat (C. Ariza)

Introducción

La Carta de Ottawa para la Promoción de la Salud¹ propugnó un nuevo enfoque de la salud pública, según el cual las intervenciones de salud pública se extienden como acciones dirigidas a mejorar el propio control de las personas para modificar los determinantes de la salud. Por otro lado, el análisis epidemiológico ha llevado con frecuencia a las intervenciones de salud pública hacia la reducción de los factores de riesgo en los individuos. De esta manera han surgido programas diferenciados para hacer frente a riesgos como el consumo de tabaco o a enfermedades como la enfermedad coronaria. Dichos programas tienden a tener sus propios objetivos, recursos y estructura de trabajo e investigación². En los últimos años han aparecido, además, programas integrados dirigidos a los determinantes del entorno físico, económico y social, más capaces de afrontar el cambio político e institucional necesario³.

Aunque nadie discute la necesidad de evaluar los resultados de las intervenciones preventivas, no hay un consenso claro en cuanto al abordaje y la metodología que deben emplearse. Así, una gran parte de las evaluaciones se centran en la calidad del diseño y de los métodos de la investigación, y menos en la calidad de la intervención (relevancia, factibilidad de la implementación...)⁴. Más allá de atender a los estándares de calidad de las intervenciones de salud pública ya establecidos para temas concretos o grupos de población⁵, es importante que las intervenciones se evalúen sistemáticamente. En este esfuerzo, son importantes los aspectos de diseño evaluativo, pero también del proceso de la intervención.

Los diseños cuasiexperimentales como mejor opción

El mejor diseño evaluativo será el que permita aproximarse al máximo a las condiciones experimentales, lo que habitualmente se conoce como «ensayos controlados aleatorizados»⁶. Éstos son el *gold standard* de los diseños de investigación. Tienen la mayor validez interna, porque son menos vulnerables a los posibles sesgos (amenazas a la validez interna), es decir, a otras explicaciones alternativas de los resultados. Son el único diseño que garantiza grupos comparables, ya que todos los posibles factores de confusión (incluso los no conocidos o no observables) se distribuyen por igual en ambos grupos⁶.

En el ámbito escolar, muchas veces no pueden implementarse los ensayos controlados aleatorizados y deben sustituirse por otros diseños. En el ámbito comunitario, a menudo es imposible aleatorizar individuos, en ocasiones por cuestiones éticas y otras veces porque no es posible evitar la contaminación entre grupos. Esto crea la necesidad de otro tipo de diseños para evaluar las intervenciones en salud pública. El diseño más común en estas circunstancias es el de los estudios orientados a examinar la plausibilidad, es decir, si los cambios observados se deben al programa o la intervención. Se conocen normalmente como diseños cuasiexperimentales y requieren extremar el control de los factores externos al estudio, lo cual puede conseguirse en gran parte con la inclusión de un grupo control o de comparación y con el ajuste por los factores confusores. Para asegurar la máxima difusión y utilidad de los resultados aportados por estos estudios, se ha elaborado la guía *Transparent Reporting of Evaluations for Non-randomized Designs* (TREND)⁷, que asegura la comunicación uniforme de los aspectos metodológicos más destacados.

Diseños evaluativos de nuevos programas en nuestro ámbito

En los últimos años hemos acumulado experiencias en la evaluación de la efectividad de intervenciones preventivas en el medio escolar, siguiendo recomendaciones como las de la guía TREND o los estándares mínimos para el diseño de un estudio de evaluación propuestos por Windsor⁸ (tabla 1).

Tabla 1

Propuesta de estándares mínimos para la evaluación de programas⁸

- | | |
|----|---|
| 1. | Diseño con grupo control y aleatorización, siempre que sea posible. |
| 2. | Representatividad de los participantes en el estudio. |
| 3. | Tamaño de la muestra, en función del efecto que quiere obtenerse (poder de 0,80, y un mínimo de 100 sujetos en cada grupo). |
| 4. | Definición completa de las variables y unos criterios mínimos de exclusión. |
| 5. | Instrumentos de autodeclaración y, si es posible, marcadores biológicos. |
| 6. | Tasa de seguimiento igual o superior al 80% en cada punto de observación. |
| 7. | Descripción completa de la intervención. |
| 8. | Evaluación del proceso para documentar la exposición a la intervención (tipo, frecuencia y duración de cada componente del programa). |

El programa de PREvención del Sida y las relaciones sexuales no protegidas en la escuela SECundària (PRESSEC)

En 1994 se inició en Barcelona la evaluación de la efectividad de un programa de prevención de las relaciones sexuales no protegidas (PRESSEC) dirigido a escolares de secundaria (15-17 años). Basado en la teoría cognitiva social, el programa promueve la modificación favorable de los conocimientos, las actitudes, las intenciones y la conducta relacionados con la transmisión del virus de la inmunodeficiencia humana (VIH). Se planteó un estudio cuasiexperimental con medidas pretest y postest, y un grupo de comparación. El programa PRESSEC, compuesto de un módulo informativo de 2 horas y otro de estrategias conductuales para la reducción del riesgo (de 5 horas), se aplicó en el grupo intervención. Para evaluar su efectividad se compararon los resultados en este grupo con los de otros dos tomados como referencia, uno al que sólo se aplicó el módulo informativo (grupo información) y otro en que no se aplicaron estrategias educativas especiales, que fue el grupo de comparación. Las escuelas de los grupos de información y de comparación podían acceder al programa en el año/curso siguiente. Las escuelas se asignaron aleatoriamente a las tres condiciones experimentales. Aceptaron participar 17 escuelas, que aportaron 220 escolares al grupo intervención, 593 al grupo de información y 402 al grupo control. En el postest, los escolares expuestos al PRESSEC fueron sexualmente menos activos, utilizaban más el preservativo cuando iniciaban las relaciones sexuales, proponían más su uso a su pareja y declaraban con más frecuencia la intención de rechazar el sexo no protegido⁹.

El programa de Prevención del Abuso de Sustancias adictivas en la Escuela (PASE)

El programa de Prevención del Abuso de Sustancias Adictivas (PASE) fue un encargo del Plan Municipal de Acción sobre Drogodependencias de Barcelona. Basado en un módulo educativo a desarrollar en lo que hoy es el primer ciclo de Educación Secundaria Obligatoria (ESO), se planteó como un programa piloto para valorar la efectividad de la intervención en comparación con un grupo de referencia, dada la inexistencia de otros programas evaluados en aquel momento en el territorio español. Las 25 escuelas de la ciudad que aceptaron participar se asignaron aleatoriamente al grupo de intervención o al grupo control, estratificando la asignación según el tipo y el tamaño de la escuela. Durante el invierno de 1990 se realizó un pretest, contestado por 2033 escolares. A continuación se administró un programa de 10 sesiones, basado en la teoría cognitiva social, con una sesión informativa en ambos grupos seguida, en el grupo de intervención, de nueve sesiones dirigidas por el docente en el aula y una sesión informativa para las familias. Tras la intervención, inmediatamente antes del verano de 1990, se realizó un postest que contestaron 1904 escolares, y 1 año después un tercer test. Aunque el programa no parecía frenar la experimentación con el tabaco, sí dis-

minuyó el paso al consumo regular en el grupo de intervención respecto al grupo de comparación¹⁰.

El desarrollo del programa PASE.bcn en el marco del European Smoking prevention Framework Approach (ESFA)

En 1998-1999 se actualizó el programa PASE, introduciendo modificaciones en el marco del proyecto europeo European Smoking prevention Framework Approach (ESFA)¹¹. Los modelos teóricos vigentes se habían ido completando a partir de la teoría de la autoeficacia de Bandura¹², el modelo de los estadios de cambio de Prochaska¹³ y la estrategia del entrenamiento de habilidades de Botvin (*Life Skill Training*, 1980)¹⁴. A la vez, el proyecto ESFA situaba las estrategias individuales de prevención en el aula en un ámbito más amplio de acciones escolares y comunitarias¹⁵. Así, el programa se actualizó como PASE.bcn, y se evaluó a medio y largo plazo¹⁶. Uno de los diez distritos de la ciudad, Sant Martí, fue escogido para el grupo de intervención. Todas las escuelas secundarias del distrito fueron invitadas a participar, y 16 de las 24 aceptaron, con participación de todas las aulas (46) de primer curso de ESO (12-13 años). El grupo control lo integraron 37 escuelas de los otros distritos de la ciudad. En cada escuela participaba un aula. Las escuelas fueron seleccionadas aleatoriamente, estratificando por el tipo de escuela (según titularidad), el tamaño de la escuela, el nivel socioeconómico del distrito de residencia de los escolares y haber participado previamente en el PASE. La estratificación resultó en nueve estratos, asegurando que en cada estrato todas las escuelas tenían la misma posibilidad de ser incluidas en la muestra. En otoño de 1998 se administró el pretest a 2015 escolares, con una tasa de respuesta del 96,9% (1083 en el grupo de intervención y 886 en el grupo control). Al inicio, las diferencias en variables sociodemográficas y consumo de tabaco u otras sustancias adictivas entre ambos grupos del estudio no fueron estadísticamente significativas. En 1999, 2000 y 2001, al final de cada año de intervención, se realizaron tres postests. Globalmente, 690 de un total de 1083 posibles escolares completaron las cuatro encuestas en el grupo de intervención y fueron seguidos hasta el final del proyecto, mientras que en el grupo control lo hicieron 603 de 886. La tasa de abandono al final del seguimiento de 3 años fue del 36,3% en el grupo de intervención y del 32,0% en el grupo control.

La nueva versión del programa se componía de siete sesiones, de las cuales dos se acompañaban de vídeos sobre influencias sociales y toma de decisiones, y otras dos contenían propuestas de entrenamiento de habilidades para rechazar la presión social y las ofertas para consumir tabaco. Se incorporó un nuevo cuaderno de trabajo para los escolares y se sometió a los docentes a dos talleres de formación. El programa se completaba con una revisión de la política del entorno escolar sobre consumo de tabaco, ofrecimiento de cesación tabáquica a los docentes e información a través de cartas y folletos informativos a las familias. Los escolares del grupo de intervención, tanto después del primer año de la intervención como al final del estudio, 3 años después, se iniciaron en menor cantidad en el consumo de tabaco que los del grupo de comparación, en ambos casos de forma estadísticamente significativa¹⁶.

El programa de prevención del consumo de cannabis en jóvenes "xkpts.com" (per què petes?)

El programa de prevención del consumo de cannabis "xkpts.com" fue diseñado en Barcelona, a partir de 2004, después de que en 5 años se doblara el consumo experimental de cannabis. Dirigido a escolares de tercer curso de ESO, esta intervención añadía dos actividades específicas de prevención del consumo de cannabis de un programa preexistente de prevención de drogodependencias, basado en la toma de decisiones ("¡Tú decides!"), evaluado anteriormente y muy bien aceptado por las escuelas de la ciudad¹⁷. Para evitar la contaminación con otras acciones de prevención de drogodependencias

más difíciles de controlar, se decidió evaluar el contenido del nuevo programa en el grupo de escuelas que el curso anterior al comienzo de la evaluación hubieran realizado el programa de prevención "¡Tú decides!". De esta manera, las 90 escuelas que cumplían esta condición fueron clasificadas en seis estratos resultantes de combinar la titularidad de la escuela (pública o concertada/privada) con tres estratos de nivel socioeconómico, derivados del índice de capacidad económica familiar¹⁸ del barrio donde estaba la escuela. A partir de ahí se asignaron escuelas de todos los estratos de forma aleatoria al grupo de intervención o al grupo control. En total 39 escuelas quedaron asignadas al grupo de intervención y otras 40 al grupo control. Para evitar al máximo la contaminación entre los grupos, y debido a que la tolerancia social del consumo de cannabis dificultaba conseguir una intervención preventiva creíble para los escolares, se decidió estudiar la evolución del consumo del grupo control 1 año antes que en el grupo de intervención, mientras se completaba el diseño de la intervención. Se trata, por tanto, de un "control histórico", en el cual los escolares pertenecen al mismo curso escolar y al mismo grupo de edad, pero son medidos para la conducta que se investiga 1 año antes que los del grupo de intervención¹⁹. En enero de 2005 se inició el pretest en el grupo control; entre enero y marzo de 2006 se realizó el pretest en el grupo de intervención, inmediatamente antes de que empezara la aplicación del programa, que duró hasta mayo de 2006. Entre abril y junio de 2006, 15 meses después del pretest, se practicó el postest en el grupo control. Finalmente, entre abril y junio de 2007, 15 meses después del pretest y 1 año después de la intervención, se realizó el postest en el grupo de intervención, para completar la recogida de datos de evaluación. Los resultados del proyecto indicaron que el programa no tenía efecto sobre el inicio en el consumo experimental de cannabis, pero sí en el paso del consumo de experimental a habitual (en los 30 últimos días)²⁰.

El proyecto de cesación tabáquica con jóvenes ASCEPIS

Este ejemplo de diseño evaluativo con grupo de control no equivalente forma parte del proyecto europeo Adolescent Smoking Cessation Project In Spain (ASCEPIS). El diseño original incluía la comparación de la efectividad de grupos terapéuticos para dejar de fumar con la de un programa informático de autoayuda ("Lo dejo ya") coordinado por un monitor o facilitador, con un tercer grupo de control o referencia. En el caso de la cesación tabáquica, para que los grupos sean totalmente comparables, deben cumplir el requisito de que todos los individuos asignados a cada grupo estén en el mismo o semejante estado de cambio de Prochaska, y, en concreto, en estado de preparación o acción¹³. Las dificultades propias en la adherencia de los jóvenes a programas de tratamiento precoz imposibilitaron conseguir un suficiente tamaño de muestra que permitiera la asignación de un grupo control. Sin embargo, a posteriori, fue posible recuperar en el postest a adolescentes que habían realizado el pretest, pero que luego no se habían incorporado a la intervención²¹. Estos individuos fueron considerados como un grupo control no equivalente: son semejantes a los intervenidos, pero el hecho de no haberse incorporado a la intervención da a entender que son diferentes, al menos en la motivación.

El control de la intervención: la evaluación del proceso

La correcta planificación de la evaluación del proceso es fundamental para obtener indicadores válidos que, en la evaluación de la efectividad de la intervención, nos permitan determinar qué estrategias han sido las más eficaces para obtener unos resultados positivos.

Indicadores para la evaluación del proceso

La evaluación del proceso es la valoración del proceso de aplicación del programa. Los componentes de la evaluación del proceso de una intervención se detallan en la tabla 2²². De todas las dimensiones

Tabla 2

Elementos que pueden formar parte de la evaluación del proceso de una intervención²²

previas, las menos y peor referenciadas son habitualmente las relacionadas con la implementación y la magnitud de la exposición a la intervención.

Hay diversos métodos para valorar la implementación, basados en variables e indicadores del desarrollo de la intervención en el aula. Aportan información sobre las estrategias, los recursos y los componentes del programa que se han aplicado, el tiempo dedicado, el número de actividades preventivas realizadas, la aceptación y la satisfacción con el programa, así como el grado de seguimiento del diseño inicial. Se han definido dos dimensiones de la implementación: cuantitativa y cualitativa²³. La dimensión cuantitativa, también conocida como cumplimiento o dosis, aporta información sobre qué cantidad de la intervención ha sido aplicada, y normalmente se expresa como el porcentaje del total del contenido que ha sido efectivamente implementado (p. ej., el porcentaje de sesiones realizadas). Por otro lado, la dimensión cualitativa o fidelidad de implementación se ha descrito como el grado en que los profesores y otros profesionales que aplican los programas los implementan tal y como señala su diseño y evaluación-piloto^{23,24}.

El cuestionario autocontestado por el profesorado (también llamado "autoinforme") y la entrevista postimplementación ("en profundidad") son métodos válidos para medir el cumplimiento, mientras que para valorar la fidelidad suele utilizarse la observación no participante de las sesiones (fidelidad observada o fidelidad objetiva), aunque en ocasiones ésta también puede valorarse a través de los cuestionarios a los participantes (fidelidad autodeclarada o fidelidad subjetiva). A modo de resumen, en la tabla 3 se muestran los diversos métodos de medición de la implementación más reconocidos en la literatura²³.

En los últimos años, gran número de estudios han señalado la importancia de evaluar la fidelidad de la implementación de las interven-

ciones preventivas escolares²⁵⁻²⁷. Su conocimiento nos puede ayudar a entender mejor el éxito o el fracaso de un programa, ya que la obtención de los resultados esperados está estrechamente vinculada a la intensidad y la calidad de la implementación. Una intervención con un buen diseño metodológico puede tener un reducido impacto si en su aplicación se produce una implementación inadecuada²⁸. De igual modo, una pobre fidelidad de implementación dificulta conocer qué componentes contribuyen al éxito de un programa, en caso de conseguir resultados positivos. No obstante, son escasos los programas preventivos que realizan una evaluación del proceso tras su aplicación, y cuando se hace, generalmente carecen del rigor necesario para garantizar la correcta aplicación de la intervención en el aula.

Resultados de la evaluación del proceso

Entre los datos que nos da la evaluación del proceso hay que diferenciar, por un lado, la consecución de los objetivos específicos del programa o intervención, y por otro los datos sobre las llamadas variables intermedias, que son aquellas que influyen y muchas veces predicen la conducta.

Resultados de la implementación de la intervención

Usualmente los docentes o quienes implementan la intervención cumplimentan un cuestionario de evaluación autocontestado (autoinforme) en el cual valoran cómo se ha desarrollado cada una de las sesiones que componen el programa. Así, el método utilizado para valorar la implementación permite medir el cumplimiento y la fidelidad autodeclarados¹⁹. En este autoinforme se valoran algunos elementos importantes para juzgar la correcta aplicación del programa. Por un lado, se valoran características relativas a la implementación de las sesiones que componen el programa, tales como el tiempo dedicado a cada una de las sesiones, las técnicas aplicadas (discusión, entrenamiento de habilidades y exposición del tema en el aula), los recursos utilizados (historietas del programa, DVD u otros elementos gráficos de la sesión didáctica audiovisual) y las actividades preventivas realizadas. Por otro, se valora el desarrollo de las diferentes sesiones en el aula, solicitando información sobre la atención de los escolares, la comprensión del contenido de la sesión por parte de los adolescentes, la facilidad para mantener la clase en calma, la satisfacción y la familiaridad con los métodos educativos, así como la satisfacción con los materiales suministrados.

En la evaluación del programa "xkpts.com", de las 117 aulas en que se aplicó se recibió el informe de evaluación de 76 (tasa de respuesta del 65%). Las 76 aulas de las cuales se recibió el informe pertenecían a 32 de las 39 escuelas participantes en el estudio¹⁹. El tiempo medio dedicado a la aplicación de la intervención en las aulas fue de 5 horas y 40 minutos, disminuyendo el tiempo dedicado a cada sesión a medida que se avanzaba en el programa. De las diversas

Tabla 3

Métodos de medición de la implementación de un programa y dimensión de su aplicación que evalúan¹⁹

Métodos de medición de la implementación	Dimensión de la implementación	¿Qué permite evaluar?
Cuestionario autocontestado por el profesorado	<ul style="list-style-type: none"> • Cumplimiento • Fidelidad (autodeclarada) 	<ul style="list-style-type: none"> • Si el profesorado ha realizado cada una de las actividades contenidas en las distintas sesiones (<i>cumplimiento</i>)
Entrevista postimplementación al profesorado	<ul style="list-style-type: none"> • Cumplimiento • Fidelidad (autodeclarada) 	<ul style="list-style-type: none"> • En qué medida el profesorado declara que ha completado cada una de las actividades principales, de acuerdo con las instrucciones propias del programa (<i>fidelidad autodeclarada</i>)
Observación no participante en el aula	<ul style="list-style-type: none"> • Fidelidad (objetiva) • Compenetración entre profesorado y alumnado a partir de la observación (Rapport) 	<ul style="list-style-type: none"> • Si el profesorado implementa cada uno de los procedimientos contenidos en las sesiones de acuerdo con la intensidad y la distribución presentes en el protocolo del programa (<i>fidelidad objetiva</i>) • Interacción del profesorado y el alumnado: <ul style="list-style-type: none"> - Capacidad del profesorado para involucrar al alumnado en las actividades y discusión - Rendimiento del profesorado en el desarrollo de determinadas estrategias

Tabla 4

Comparación de las variables de la evaluación del programa PRESSEC de prevención de las relaciones sexuales no protegidas. Diferencias al inicio y al final del estudio, según grupos (Barcelona, 1994)^a

	Control			Información			Programa PRESSEC		
	Inicio	Final	p	Inicio	Final	p	Inicio	Final	p
<i>Conocimientos (puntuación media)^a</i>									
Sida	8.67	8.62	NS	8.40	9.02	<0.001	8.58	9.06	<0.001
Embarazo	8.58	8.84	NS	8.34	9.12	<0.001	8.67	9.45	<0.01
Contracepción	8.76	9.09	<0.01	8.77	9.05	<0.01	8.87	9.30	<0.01
<i>Actitudes (puntuación media)^a</i>									
Uso del preservativo	9.16	9.31	NS	9.25	9.43	<0.05	9.12	9.51	<0.001
Susceptibilidad percibida	9.17	8.99	NS	9.21	9.21	NS	9.50	9.40	NS
<i>Intención de conducta (puntuación media)^a</i>									
Uso del preservativo	8.60	8.93	NS	8.86	9.20	<0.05	9.18	9.54	NS
Rechazo de las relaciones sexuales protegidas	4.77	4.75	NS	4.63	5.33	<0.01	4.74	5.47	<0.05
<i>Conducta declarada (%)</i>									
Sexualidad activa	28%	33%	NS	24%	27%	NS	26%	31%	NS
Uso habitual del preservativo	26%	28%	NS	21%	22%	NS	22%	26%	<0.05

^a Puntuación: 0-10.

técnicas propuestas en el protocolo del programa para el desarrollo de las sesiones, fue la discusión la que se aplicó con mayor frecuencia, ya que en el 100% de las aulas se discutió sobre el tema en algún momento de la intervención. El entrenamiento de habilidades se aplicó en el 59,2% de las aulas en la tercera sesión, tal y como especificaba el protocolo, si bien un 13,2% lo aplicaron en una sesión distinta. De los recursos facilitados, el DVD fue el más utilizado y se proyectó en el 93,4% de las aulas (aunque el 9,2% lo utilizaron en una sesión diferente a la indicada según el protocolo). En cuanto a las actividades preventivas, la media de actividades realizadas fue de 9,2 de las 16 propuestas, siendo en la segunda y la tercera sesiones cuando se realizó una mayor proporción de actividades, 2,2 de 3 y 2,2 de 3, respectivamente. De manera global, entre las actividades propuestas, el visionado del DVD y la posterior discusión fueron las realizadas con mayor frecuencia (84,2% de las aulas).

Otro de los objetivos de la evaluación del proceso es la evaluación de la implantación y de la diseminación del programa. Tras la evaluación de la prueba piloto del programa PASE de Barcelona, se intentó su implantación generalizada y en 2 años se llegó a 80 escuelas (17,9% de todas las de Barcelona). Sin embargo, la aplicación no fue uniforme en toda la ciudad, con una amplia variación de unos distritos a otros. Un análisis más detallado de este hecho reveló que la actuación del agente intermediario, en este caso el equipo de salud del distrito, era clave en la aplicación del programa. En los distritos en que los programas de educación de la salud no eran una prioridad o en aquellos donde no había un equipo estable de salud, la extensión del programa era muy inferior.

Resultados en las variables intermedias

Se puede discutir si es una evaluación de resultados propiamente dicha, ya que no son la variable conductual principal que queremos modificar. Sin embargo, de acuerdo con los modelos teóricos actualmente más aceptados, sabemos que algunas variables, como el nivel de conocimientos, las creencias, los valores y las actitudes, las expectativas de autoeficacia y la autoestima, median en la adquisición de la conducta.

En el programa PRESSEC, prácticamente todas las variables intermedias (los conocimientos sobre el sida, el embarazo y la contracepción, las creencias sobre el uso del preservativo y las intenciones de rechazar el sexo no protegido) mostraron cambios estadísticamente

significativos en el grupo de intervención respecto al grupo de comparación⁹ (tabla 4).

El conjunto de los indicadores de la evaluación del proceso y los indicadores de las variables intermedias, también llamados indicadores intermedios, se conoce como la "caja negra" (*black box*) de la intervención³⁰, ya que permite explicar los mecanismos por los que actúa la intervención, pero su contribución a los resultados es en general poco visible y escasamente conocida.

Grado de cumplimiento de la intervención

Una cuestión relevante de la evaluación del proceso es poder conocer qué escuelas están completando adecuadamente la intervención. Por este motivo, es recomendable establecer indicadores que puedan definir el grado aceptable de cumplimiento de una intervención.

En la evaluación del programa PASE.bcn, dentro del marco de ESFA, de las 13 escuelas que completaron el proyecto, tres lo implementaron adecuadamente el primer año, irregularmente el segundo y prácticamente nada en el tercero. En este caso se consideró «estar expuesto a la intervención» haber completado al menos el 50% de las actividades planificadas en el protocolo (intervención aceptable). De acuerdo con ello, 10 de las 13 escuelas fueron incorporadas al análisis para examinar el efecto de la intervención. Independientemente de que el análisis global del estudio debe investigar si una intervención que no es desarrollada por todas las escuelas es replicable más allá de la fase de evaluación inicial, lo que está fuera de duda es que no es posible esperar efecto de un programa que no se ha desarrollado adecuadamente. En el caso que nos ocupa, si incluíamos en el análisis todas las escuelas, los efectos de la intervención eran estadísticamente significativos después del primer año (cuando se tenía constancia de que todas las escuelas habían implementado correctamente la intervención), pero el efecto desaparecía en el análisis a los 3 años de la intervención. Contrariamente, si se excluía a las escuelas con un bajo grado de cumplimiento, el efecto se mantenía también a los 3 años¹⁶.

En la evaluación del programa de prevención del consumo de cannabis "xkpts.com", para valorar de forma global la implementación, y de acuerdo con el buen rendimiento conseguido en el estudio ESFA, se definieron operativamente dos indicadores de cumplimiento. Por un lado, siguiendo un criterio cuantitativo a la hora de aplicar las actividades en el aula, se definió como intervención correcta o acep-

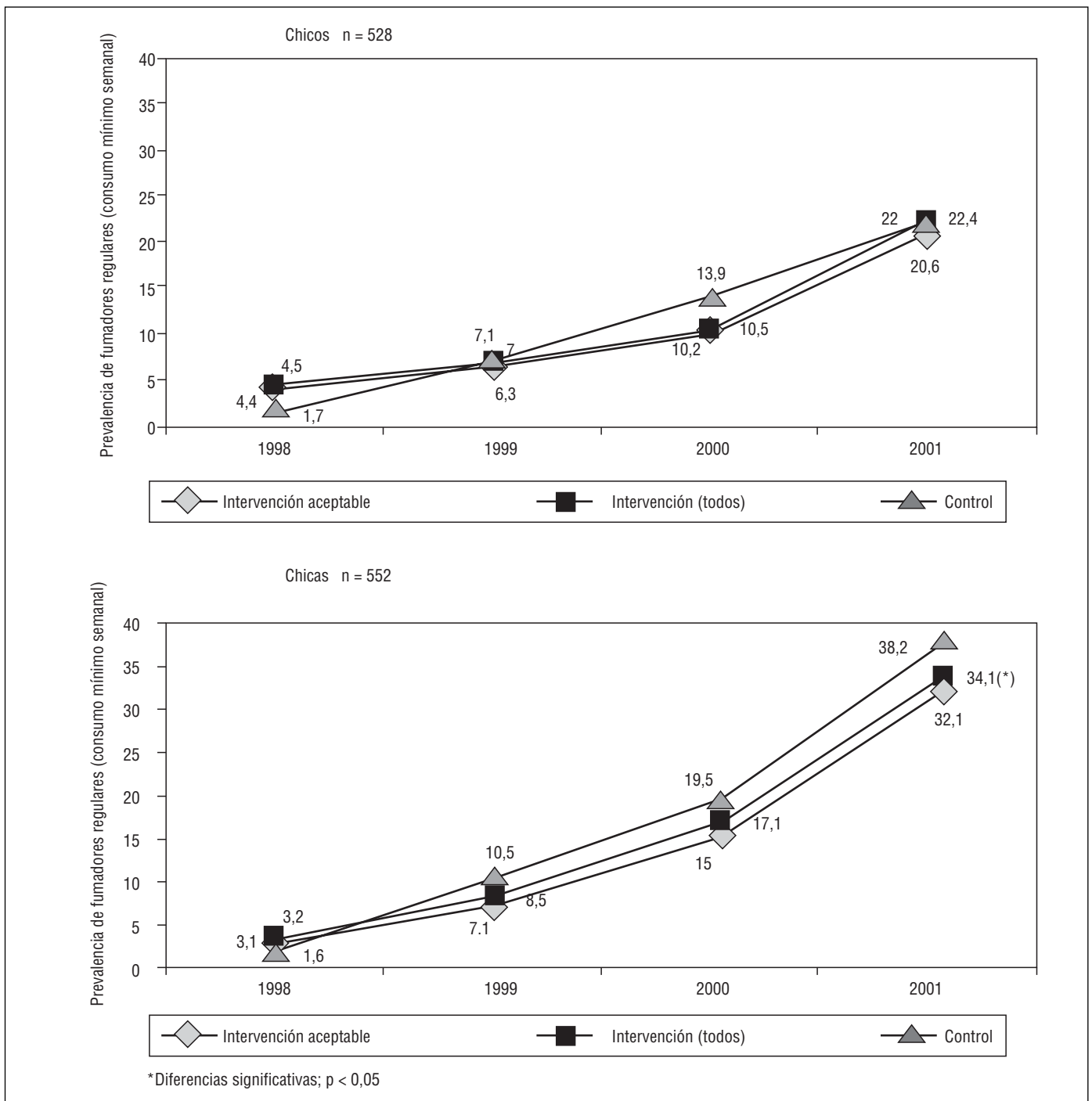


Figura 1. Evolución de la prevalencia del tabaquismo en escolares, según la aplicación del programa PASE.bcn dentro del Proyecto ESFA. Barcelona, 1998-2001.

table aquella en que se habían realizado al menos ocho de las 16 actividades preventivas previstas. Por otro lado, siguiendo un criterio cualitativo, se definió como intervención cualificada aquella en que se había observado de forma rigurosa el protocolo de la intervención, realizando un mínimo de una actividad para cada una de las sesiones generales y dos actividades para cada una de las sesiones específicas. Se observó que 45 aulas de las 76 (59,2%) habían realizado la intervención de forma aceptable, mientras que sólo 29 de las 76 (38,2%) habían realizado una intervención cualificada y de acuerdo con el protocolo¹⁹. Estos datos concuerdan con la evidencia científica que muestra la dificultad de conseguir un cumplimiento adecuado en las intervenciones preventivas escolares^{24,28}.

El último de los elementos importantes a tener en cuenta en la evaluación del proceso es la consideración del contexto. Es importante que las intervenciones preventivas escolares puedan contar con actividades y políticas de entorno que refuercen el mensaje que transmiten las intervenciones individuales en el aula. Por ejemplo, en el caso del proyecto PASE.bcn, la intervención activa y guiada en el entorno escolar incrementó el proceso de adopción de cambios. Así, orientar a los centros sobre la regulación del consumo de tabaco, facilitar guías para la implementación de cambios y ofrecer cesación al personal escolar se revelaron como medidas complementarias clave para cambiar el marco de la regulación del consumo de tabaco en la escuela³¹.

La interpretación de los resultados: la evaluación del efecto

El establecimiento de la efectividad de una intervención a menudo suele incluir la evaluación del impacto y de los resultados. Aunque los términos "impacto" y "resultados" se utilizan muchas veces con significados distintos, para algunos autores la evaluación del impacto examina el efecto inicial del programa en los objetivos de cambio más próximos, como las políticas, las conductas y las actitudes. De esta manera, la evaluación del impacto se correspondería con la valoración de los objetivos iniciales de la intervención²². Con esta perspectiva, la evaluación de los resultados se refiere al efecto consecuente de un programa o intervención en los resultados de salud de una población, en correspondencia con la finalidad u objetivo principal del programa. La evaluación de resultados se ha llamado con frecuencia «evaluación sumativa», porque el evaluador está en una posición favorable para hacer una consideración global sobre el valor del programa. Esta consideración asume el éxito previo de la evaluación del proceso y del impacto²².

Relación entre la evaluación del proceso y la evaluación de la efectividad

Una cuestión de particular interés en la interpretación de los resultados de la evaluación de una intervención preventiva es la relación entre la evaluación del proceso y la evaluación del efecto o de los resultados. Es frecuente que en la planificación de las intervenciones se expliciten relativamente bien las actividades previstas; poco, en cambio, se suele decir respecto a cómo se ha desarrollado la intervención prevista, qué se ha llegado a hacer (o no) y cómo se considera que ha funcionado. La descripción precisa de todos estos aspectos propicia la evaluación del proceso, que con frecuencia es olvidada en estudios muy centrados en la medición de resultados de salud. Poco podemos saber de por qué una determinada intervención no ha funcionado si no tenemos información de lo que se ha realizado. Nos parecería raro evaluar el efecto de un fármaco que sólo se ha administrado de forma errática, sin cumplir la posología recomendada; sin embargo, esto no es infrecuente en promoción de la salud. Además, hay que tener en cuenta la gran complejidad de estas intervenciones, en comparación con la simple administración de un fármaco.

Cuando se dispone de información y nos podemos pronunciar sobre el nivel de realización de la intervención, es útil establecer un grado de aceptabilidad de la intervención. Ello posibilitará una doble comparación: intentar saber cuáles son los resultados en aquellos que han realizado la intervención de forma correcta respecto al grupo de comparación (sobre los que no se ha desarrollado el programa), y también saber si hay diferencias entre estos últimos y los que han estado expuestos parcialmente o de manera insuficiente a la intervención. En la figura 1 se muestra la evolución del consumo regular de tabaco a lo largo de los 4 años de duración del proyecto ESFA de evaluación del programa PASE.bcn¹⁶, diferenciando por sexo. De entrada, puede observarse que para ambos sexos el consumo regular de tabaco (como mínimo semanal) era menor en los escolares del grupo control. Estas cuestiones son difíciles de obviar en salud comunitaria. En este caso, las diferencias no fueron significativas, pero podrían ir asociadas a que el distrito de la ciudad donde se emplazó el grupo de intervención tenía unos indicadores de nivel socioeconómico más desfavorables que el resto de la ciudad. Al cabo de 1 año e inmediatamente después de haberse aplicado el programa en el aula, la evolución del consumo regular del grupo control fue claramente superior en el grupo control en ambos sexos respecto al grupo de intervención. Entre los chicos, este consumo creció un 6% en el grupo control, mientras que fue de tan sólo un 1,9% en los que realizaron la intervención aceptablemente. En las chicas, estas diferencias fueron del 8,9% y el 4%, respectivamente. Pero ya después del primer año era posible comprobar las diferencias entre quienes habían realizado mejor la intervención y el conjunto de todos los intervenidos. Con

excepción del segundo año de intervención para los chicos, a lo largo del proyecto se siguieron observando diferencias en el grupo de intervención, con mejores datos en el subgrupo que recibió la intervención con una calidad aceptable. De este hecho se deducen dos mensajes importantes a diseminar tanto entre los docentes que implementan los programas como entre los propios escolares, que se ilustran en la figura 1 en relación con la evaluación del programa PASE.bcn: en primer lugar, que una exposición mínima a la intervención puede tener un efecto (en la figura se ve en las chicas en los años 2000 y 2001); y en segundo lugar, que hacer bien la intervención tiene un efecto añadido (en la figura es visible en ambos sexos con los datos de 2001).

Otros estudios han reflejado resultados comparables. El estudio de Wiborg et al³² para evaluar el *Smoke-free Class Competition* en Alemania permitió comprobar que cuando se comparaban los alumnos que no habían realizado el programa con los que habían estado expuestos a él, sin haberlo completado (por transgredir las normas antes del final), había diferencias en el consumo. No obstante, estas diferencias en el consumo eran mayores cuando se comparaba a los controles con los que habían mantenido el compromiso de no fumar durante los 6 meses de intervención.

En conclusión, aquellos grupos de escolares expuestos a un nivel aceptable de una intervención suelen presentar posteriormente resultados mejores en la conducta que se quiere prevenir, respecto a los individuos expuestos con un cumplimiento insuficiente de la intervención. Ello demanda una adecuada planificación de la evaluación del proceso de la intervención, así como una documentación e información precisa tanto en informes como en publicaciones científicas de la evaluación de dicha intervención.

Contribuciones de autoría

C. Ariza ha realizado el borrador y la versión final. Todos los autores han participado en la discusión y revisión del manuscrito.

Financiación

Este artículo ha sido elaborado con el apoyo del Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya (AGAUR SGR 2009-1345).

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

1. WHO (World Health Organization), Health and Welfare Canada, Canadian Public Health Association. Ottawa Charter for Health promotion. WHO Regional Office for Europe. Copenhagen; 1986.
2. Nutbeam D. Achieving "best practice" in health promotion: improving the fit between research and practice. *Health Educ Res.* 1996;11:317-25.
3. Nutbeam D. Evaluating health promotion-progress, problems and solutions. *Health Promotion Int.* 1998;13:27-44.
4. Speller V, Learmonth A, Harrison D. The search for evidence of effective health promotion. *BMJ.* 1997;315:361-3.
5. Task Force on Community Preventive Services. Introducing the Guide to Community Preventive Services: methods, first recommendations and expert commentary. *Am J Prev Med.* 2000;18(1 Suppl):1-142.
6. West SG, Duan N, Pequegnat W, et al. Alternatives to the randomized controlled trial. *Am J Public Health.* 2008;98:1359-66.
7. Des Jarlais DC, Lyles C, Crepaz N, and the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health.* 2004;94:361-6.
8. Windsor R, Baranowski T, Clark N, et al. Evaluation of health promotion, health education, and disease prevention programs. Mountain View, California: Mayfield Publishing Company; 1994.
9. Diez E, Juárez O, Nebot M, et al. Effects on attitudes, knowledge, intentions and behaviour of an AIDS prevention programme targeting secondary school adolescents. *Promotion & Education.* 2000;7:17-22.

10. Villalbí JR, Aubá J, García-González A. Evaluación de un proyecto de prevención primaria del tabaquismo: el proyecto piloto PASE de Barcelona. *Rev San Hig Pub.* 1992;66:143-8.
11. De Vries H, Mudde A, Leijts I, et al. The European Smoking prevention Framework Approach (ESFA): an example of integral prevention. *Health Educ Res.* 2003;18:611-26.
12. Bandura A. *Social foundations of thought and action: a social cognitive theory.* New York: Prentice Hall; 1986.
13. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol.* 1983;51:390-404.
14. Botvin GJ, Eng A, Williams CL. Preventing the onset of cigarette smoking through life skills training. *Prev Med.* 1980;9:135-43.
15. Ariza C, Nebot M, Jané M, et al. El proyecto ESFA en Barcelona: un programa comunitario de prevención del tabaquismo en jóvenes. *Prev Tab.* 2001;3:70-7.
16. Ariza C, Nebot M, Giménez E, et al. Longitudinal effects of the European Smoking prevention Framework Approach (ESFA) Project in Spanish adolescents. *Eur J Public Health.* 2008;18:491-7.
17. Calafat A, Amengual M, Mejias G, et al. Evaluación del programa de prevención 'Tú decides'. *Adicciones.* 1989;1:96-111.
18. Ventura A, Càrcel C, Canals RM, et al. Index de capacitat econòmica familiar II [Family economic capacity index II]. Barcelona: Gabinet de Programació. Publicacions de l'Ajuntament de Barcelona; 1999.
19. Sánchez-Martínez F, Ariza C, Pérez A, et al. Evaluación de proceso del programa escolar de prevención del consumo de cannabis en adolescentes "xkpts.com". *Adicciones.* 2010;22:217-26.
20. Ariza C, Pérez A, Sánchez-Martínez F, et al. Efectividad de una intervención preventiva sobre el consumo habitual de cannabis en escolares de Barcelona. *Gac Sanit.* 2008;22:133.
21. Ariza C, González I, Juárez O, et al., y Equipo de Coordinadores del Proyecto "Adolescent Smoking Cessation" en España. Evaluación de un programa de cesación en escolares del proyecto "Adolescent Smoking Cessation (ASC)" en España. *Prev Tab.* 2005;7(Supl 1):139.
22. Rychetnik L, Hawe P, Waters E, et al. A glossary for evidence based public health. *J Epidemiol Community Health.* 2004;58:538-45.
23. Resnicow K, Davis M, Smith M, et al. How best to measure implementation of school health curricula: a comparison of three measures. *Health Educ Res.* 1998;13:239-50.
24. Dusenbury L, Brannigan R, Hansen WB, et al. Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. *Health Educ Res.* 2005;20:308-13.
25. Payne AA, Gottfredson DC, Gottfredson GD. School predictors of the intensity of implementation of school-based prevention programs: results from a national study. *Prev Sci.* 2006;7:225-37.
26. Pankratz MM, Jackson-Newsom J, Giles SM, et al. Implementation fidelity in a teacher-led alcohol use prevention curriculum. *J Drug Educ.* 2006;36:317-33.
27. Sánchez V, Steckler A, Nitriat P, et al. Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Educ Res.* 2007;22:95-107.
28. Resnicow K, Botvin G. School-based substance use prevention programs: why do effects decay? *Prev Med.* 1993;22:484-90.
29. Villalbí JR, Ballestín M, Nebot M, et al. The prevention of substance abuse in schools: a process evaluation of the adoption of a standardised education module. *Promotion and Education.* 1997;4:15-9.
30. Macdonald G, Veen C, Tones K. Evidence for success in health promotion: suggestions for improvement. *Health Educ Res.* 1996;11: 367-76.
31. Valmayor S, Ariza C, Tomás Z, et al. Evaluación de una intervención de prevención del tabaquismo en el entorno escolar. *Prev Tab.* 2004; 6:18-25.
32. Wiborg, G, Hanewinkel, R. Effectiveness of the "Smoke-Free Class Competition" in delaying the onset of smoking in adolescence. *Prev Med.* 2002;35:241-9.