

Francisco Alvira Martín

Metodología de la evaluación de programas: un enfoque práctico



Editorial LUMEN / HVMANITAS
Viamonte 1674

1055 Buenos Aires

☒ 373-1414 (líneas rotativas) Fax (54-1) 375-0453
República Argentina

Colección Política, servicios y trabajo social

Directores: Dr. **Ezequiel** Ander-Egg
Dra. María José **Aguilar Idáñez**

Director editorial **LUMEN/HVMANTAS**: Dr. Ezequiel Ander-Egg

Revisión de texto: Pablo Valle

Coordinación gráfica: Lorenzo **Ficarelli**

Diseño de tapa: **Oscar Sánchez** Rocha

Diseño Ilustración: **María** José Aguilar Idáñez

ISBN 950-724-603-7

© 1997 by LUMEN / HVMANTAS

Hecho el depósito que previene la ley 11.723

Todos los derechos reservados

LIBRO DE EDICIÓN ARGENTINA

PRINTED IN ARGENTINA

Índice

| | |
|---|-----|
| 1. Introducción: aspectos generales del proceso de evaluación | 7 |
| 2. La familiarización con el programa | 19 |
| 1. Los procedimientos | 21 |
| 2. Técnicas y ejemplos | 27 |
| 3. El diseño de la evaluación | 33 |
| 1. La determinación de objetivos | 35 |
| 2. El diseño propiamente dicho | 46 |
| 4. La recogida de los datos | 67 |
| 1. Ejemplo 1..... | 76 |
| 2. Ejemplo 2..... | 90 |
| 5. Análisis e interpretación de resultados | 103 |
| 1. El análisis de tablas de contingencia..... | 107 |
| 2. El análisis multivariante | 111 |
| 6. Conclusión | 123 |
| Anexos metodológicos | 127 |
| 1. La entrevista focalizada..... | 129 |
| 2. El grupo de discusión / La reunión de grupo | 138 |
| 3. La encuesta | 143 |
| 4. La observación sistemática..... | 150 |
| 5. El análisis discriminante..... | 156 |
| 6. El análisis de series temporales..... | 165 |
| Bibliografía citada..... | 173 |

del programa deberían ir encaminadas a modificar los estereotipos y romper la cultura empresarial "machista", siguiendo el mecanismo de acciones positivas en la empresa mediante convenios voluntarios. El análisis documental **mostró** las inconsistencias existentes pero sólo las entrevistas confirmaron la existencia de dos modelos de intervención, lo que daba lugar a actividades y objetivos distintos y, en el fondo, a programas diferentes. Éste, se vio más tarde, era uno de los problemas básicos del programa CAPEM desde el punto de vista de su evaluación.

En este ejemplo, las reflexiones que se realizan se hacen desde el punto de vista del evaluador externo, puesto que el Instituto de la Mujer contrató dicha evaluación externa con el autor del libro y éstas son sus reflexiones. Sin embargo, es evidente que dentro del Instituto existía preocupación por el funcionamiento del programa; una evaluación interna hubiera llegado a los mismos resultados, pero era difícil de realizar por:

- la falta de técnicos para realizarla, y
- la **fuerte** implicación del propio Instituto en el programa.

El diseño de la evaluación

1. La determinación de objetivos
2. El diseño propiamente dicho
 - a. La evaluación de necesidades
 - b. La evaluación del diseño y la conceptualización de la intervención
 - c. La evaluación de la implementación
 - d. La evaluación de resultados y de impacto
 - e. La evaluación económica

Diseñar una evaluación es plantear qué información se va a recoger y cómo se va a recoger y analizar teniendo siempre a la vista los objetivos. La lógica que conecta información, método de recogida de información, análisis y objetivos de la evaluación es el diseño de la misma. Esta lógica, este diseño, debe ser adecuado para los objetivos que se persiguen, de modo que la conexión señalada (información-métodos-análisis-objetivos) esté bien hecha y no se preste a interpretaciones contrapuestas.

En esta fase, es conveniente separar dos procesos íntimamente entrelazados:

- La determinación de los objetivos de la evaluación.
- El diseño propiamente dicho.

Resulta evidente que la fijación de objetivos condiciona el diseño, a la vez que es una condición necesaria; sin objetivos de evaluación, es imposible pensar en un diseño concreto.

1. La **determinación** de objetivos

La fijación de objetivos condiciona - como se ha dicho — el diseño, pero también el tipo de evaluación que se va a efectuar y qué criterios de valor se van a utilizar. Resulta inexacto señalar que un aspecto condiciona otros, dado que es un condicionamiento mutuo. La imagen podría ser la de una pirámide de cuatro lados (ver figura 1) en la que la base serían las preguntas que sobre el programa formula el "cliente" y a las que la evaluación debe dar respuestas adecuadas. Cada uno de los cuatro lados —**criterios** de valor, objetivos, tipo de evaluación y **diseño**— se apoyaría en las preguntas (la base) y estaría estrechamente interconectado con los otros lados de la pirámide.

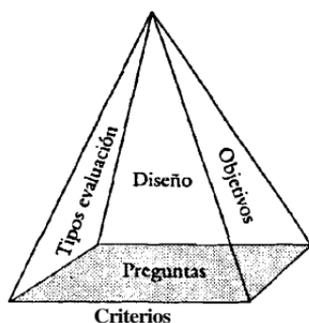


Figura 1

Desde un punto de vista analítico y secuencial, el punto de partida lo constituyen **las preguntas**.

Las evaluaciones persiguen diferentes objetivos. Dejando a un lado la distinción enue evaluación formativa (cuyo objetivo último es modificar el programa para mejorarlo o adaptarlo mejor) y evaluación sumativa (cuyo objetivo último es dar información para decidir sobre la continuación o terminación de un programa), **el objetivo general de una evaluación es proporcionar información para dar respuesta a las preguntas que se plantean sobre los programas.**

Por eso, la mejor manera de centrar los objetivos de una evaluación consiste en formular preguntas relativas,al programa que se quieran ver contestadas en la evaluación. En el cuadro adjunto, se puede ver un listado de preguntas de evaluación pero ¿quién formula estas preguntas? y ¿cómo se llega a ellas?

Las preguntas de la evaluación según tipos

Evaluación de necesidades

- ¿Se ha elaborado un inventario de recursos necesarios y disponibles?

- ¿Cuál es el alcance del problema y la distribución de la población objeto o potencialmente usuaria?
- ¿Cuál es el volumen del problema y dónde se encuentra?

Evaluación del diseño

- ¿Existe una coherencia, en el programa estudiado, entre análisis de la realidad, objetivos generales, objetivos específicos, actividades y metodología?
- ¿Responden los objetivos a las necesidades detectadas?
- ¿Son congruentes los objetivos con la "filosofía" de la organización o grupo que los formula?
- ¿Están los objetivos específicos bien formulados con respecto a los objetivos más generales?
- ¿Están los objetivos correctamente jerarquizados?, ¿responden esas prioridades a las necesidades reales?
- ¿Se corresponden las actividades a la jerarquización de objetivos?
- ¿Existen o se prevén los recursos necesarios?
- ¿Está prevista una metodología de intervención: secuenciación de tareas, calendario y temporización, mecanismos de control y retroalimentación, etc.?
- ¿Se han establecido con claridad la naturaleza y la magnitud de los problemas y necesidades?
- ¿Se encuentra adecuadamente conceptualizado el problema social?
- ¿Funcionará un programa de intervención determinado en condiciones óptimas?
- ¿Subyace una base lógica en el programa?
- ¿Existe coherencia entre el programa elaborado y los resultados del diagnóstico?, ¿responde el programa a las necesidades y problemas detectados durante el diagnóstico?
- ¿Se ha especificado un calendario de ejecución?
- ¿Se han previsto recursos suficientes, y en tiempo oportuno, para alcanzar los objetivos y metas propuestas?

- ¿Se han realizado una adecuada estructura financiera y un calendario financiero del programa?
- ¿Se ha previsto una estructura de gestión del programa?

Evaluación de resultados

- ¿Cuáles son los productos del programa?, ¿se han dado en la magnitud que estaba prevista?
- ¿Se han dado productos inesperados?
- ¿Es efectivo el programa en la consecución de los objetivos que se persiguen?
- ¿Cómo se ha modificado la situación de los usuarios con el programa?
- ¿Se pueden deber los resultados del programa a procesos alternativos exógenos al mismo?
- ¿Qué resultados o efectos no previstos han surgido como consecuencia del programa?
- ¿Se están obteniendo los resultados esperados?
- ¿Cuáles han sido los efectos del programa?

Evaluación de impacto

- ¿Cuáles han sido los efectos del programa?
- ¿Cuál ha sido el impacto en los medios de comunicación?
- ¿Cuál ha sido el impacto sobre la comunidad y poblaciones no objeto del programa?

Evaluación del proceso y funcionamiento del programa

- ¿Se ejecuta el programa o actividad como se pretendía?
- ¿Se llega al grupo de personas previsto?
- ¿Se benefician todos los destinatarios de igual manera?
- ¿Qué grado de participación y satisfacción se consigue?
- ¿Existe una demanda?, ¿se cubre?, ¿cómo y por qué fluctúa?, etc.

- ¿La importancia relativa concedida a cada actividad o sesión es la apropiada?
- ¿Existe demanda no cubierta?
- ¿Quiénes y cuántos son los usuarios del programa?
- ¿Los métodos elegidos son empleados con la máxima eficacia?, etc.
- ¿Existen los recursos necesarios?
- ¿Se utilizan los recursos —económicos, materiales, humanos— de una manera adecuada?, ¿se pueden explotar u optimizar mejor?, etc.
- ¿Funciona el equipo adecuadamente?, ¿cómo se funciona en las reuniones?, ¿son productivas?
- ¿Cómo se dan los flujos de información?
- ¿Están bien organizadas y distribuidas las tareas?
- ¿Asume cada uno su responsabilidad?
- ¿Existen solapamientos o lagunas en la distribución de funciones?
- ¿Existen conflictos?, ¿se resuelven de manera constructiva?
- ¿Cómo se interacciona con los participantes?, etc.
- ¿Se implementa el programa como estaba previsto en el diseño?
- ¿El servicio que se ofrece tiene la calidad necesaria?

La respuesta es obvia: las preguntas las formula el cliente o, mejor, el evaluador a través del cliente, entendiendo como cliente la persona o personas que impulsan la evaluación, la patrocinan o encargan. Una variante del "cliente" son las "partes interesadas", es decir, las diferentes personas implicadas en un programa y que pueden estar interesadas en su evaluación (políticos, directivos, técnicos/profesionales, usuarios, familias de usuarios, comunidad en general...).

El procedimiento normal para especificar las preguntas que se quieren ver contestadas en la evaluación es la entrevista abierta con el "cliente"; en dicha entrevista, deben precisarse:

- Las razones para hacer la evaluación.
- La información que se necesita sobre el programa.
- La utilidad de la evaluación, etc.

Para llegar a todas estas precisiones, lo más fácil es conseguir que el "cliente" formule en voz alta las preguntas sobre el programa que se debe evaluar.

Puede que existan varias preguntas, y no una sola, sobre un programa, pero, o bien podrán ordenarse en torno a un tipo común de preguntas, o bien existirá una sola que es la esencial y servirá de guía-base. Al final, el evaluador dispondrá de una o varias preguntas similares a las que aparecen en el cuadro anterior.

Las preguntas en la evaluación de la ONCE

En los años 1991 y 1992, el autor del libro llevó a cabo una asesoría externa en la ONCE (Organización Nacional de Ciegos de España) para el diseño e instrumentación de un sistema de información para la evaluación de los servicios y programas para afiliados.

Uno de los primeros pasos consistió en centrar los objetivos de dicho sistema de información desde el punto de vista de la evaluación.

Para ello, se procedió a entrevistar al director del Servicio para Afiliados y a una muestra del Consejo Político de la Organización, Consejo que representa a los afiliados. Como era de esperar, dirección técnica y Consejo Político tenían intereses no contrapuestos pero sí diferentes; sin embargo, las entrevistas abiertas permitieron centrar la atención en las preguntas que realmente preocupaban a ambos niveles, niveles que luego, más adelante en el proceso, se complementan con el nivel técnico y profesional. Las preguntas que surgieron fueron las siguientes:

- ¿Están atendidas todas las necesidades de los afiliados con los programas y servicios existentes? ¿Queda algo por cubrir?
- ¿La información sobre programas, prestaciones y servicios llega a todos los afiliados?
- ¿Cuál es la calidad de los programas y servicios prestados?
- ¿Cómo se controla la labor de los profesionales?
- ¿Se emplean bien los recursos humanos y económicos? ¿Sobran o faltan?

Formuladas las preguntas, el evaluador tiene que transformarlas en **objetivos de evaluación** precisando, a la vez, el **tipo de evaluación** y los **criterios de valor** que se van a utilizar. Esta concreción del tipo de evaluación y criterios de valor condiciona, casi determina, el **plan/diseño de la evaluación**.

El esquema adjunto refleja los **tipos de evaluación** más frecuentes. Sin embargo, es muy difícil incluir todos los posibles tipos de evaluación, dado que existen más de treinta, muchos de los cuales son sólo diferentes en la utilización de distintas etiquetas o nombres. Así, la evaluación de la efectividad es lo mismo que la evaluación de la eficacia, y ambas coinciden con la evaluación de resultados; la evaluación del proceso realmente cae dentro de la evaluación de la instrumentación (o aplicación) y del funcionamiento del programa, y así sucesivamente.

Tipos de evaluación

1. Evaluación de necesidades.
2. Evaluación del diseño y conceptualización del programa.
3. Evaluación de la implementación/aplicación.
 - 3.a Evaluación de cobertura.
 - 3.b Evaluación del funcionamiento.
4. Evaluación de resultados.
5. Evaluación de impacto.
6. Evaluación económica.

El valor que tiene este esquema es que sigue un orden lógico y jerárquico, y hace hincapié en los tipos más frecuentes e importantes.

Un programa/intervención surge ante un problema que se debe resolver; de ahí que la primera tarea sea la de identificar exacta-

mente el problema, estudiando y analizando sus características, y avanzando, incluso, soluciones. En esto consiste, muy resumidamente, una evaluación de necesidades.

La elaboración del programa parte —o debe partir— de los resultados de la evaluación de necesidades, y se apoya en los conocimientos básicos, disponibles en las diferentes ciencias sociales, del problema y de su explicación. Elaborado el programa a nivel documental y **teórico**, lo normal es realizar una evaluación **de su conceptualización** y diseño. Ésta **es** una evaluación sobre el papel y busca la coherencia lógica entre los resultados de la evaluación de necesidades, el programa y los conocimientos básicos sobre el problema por resolver; este tipo de evaluación puede descubrir una mala traducción del modelo teórico causal, una insuficiente previsión de recursos, una inadecuada previsión de la prestación del servicio para llegar a la población objeto, etc.

Una vez asegurada la coherencia del programa elaborado, lo habitual es ponerlo en marcha; esta aplicación puede también ser evaluada centrándola:

- bien en la **cobertura** del mismo,
- bien en su funcionamiento en relación con lo previsto.

Con el programa funcionando según lo previsto, lo más adecuado es la evaluación **de** resultados. Ésta se centra en estudiar los efectos que tiene sobre la población objeto, distinguiendo los efectos que se le pueden atribuir de los no atribuibles, los efectos perseguidos de los no queridos...

Cuando la evaluación se centra en los efectos del programa sobre la comunidad, es decir, más allá de la población objeto, nos encontramos con la evaluación **de** impacto.

Por **último**, el análisis de la eficiencia **del** programa (evaluación coste-beneficio, coste-efectividad o coste-utilidad) cierra el círculo al introducir el criterio de recursos escasos en la evaluación para buscar alternativas efectivas y eficientes.

Los tipos de evaluación en la ONCE

Las preguntas reseñadas en la evaluación de la ONCE se centran todas en objetivos de evaluación presentes en evaluaciones de cobertura, de funcionamiento del programa y de eficiencia.

La evaluación de cobertura recoge la atención a los afiliados y su información; la evaluación del funcionamiento recoge las preguntas sobre la calidad y control de los profesionales, y la de eficiencia recoge la última pregunta.

Resulta interesante constatar aquí cómo la evaluación de resultados queda fuera de las preocupaciones del Consejo Político y de la Dirección Técnica de la organización.

Los **criterios de valor**, necesarios para efectuar el correspondiente juicio de valor, están estrechamente relacionados con los tipos de evaluación, de modo que surgen y se explicitan al centrar la evaluación en un tipo específico. Así, en la evaluación económica, el criterio de valor es la eficiencia, es decir, la consecución de unos efectos con el menor coste posible; si existen alternativas menos costosas, la evaluación económica será negativa. Del mismo modo, en la evaluación de resultados, el criterio es la consecución de los efectos perseguidos o, en todo caso, de efectos positivos; si el programa no produce los efectos deseados o efectos positivos, la evaluación será negativa.

De igual modo que el **tipo** de evaluación que se va a efectuar se sigue de las preguntas y objetivos establecidos, los criterios de valor surgen explícita o implícitamente de preguntas, objetivos y tipo de evaluación. El cumplimiento de los objetivos del programa es el criterio de valor más utilizado, pero existen otros que especifica el "cliente" de la evaluación y/o las "partes interesadas".

Cuando se utilizan como criterios de valor los objetivos del programa, el procedimiento es directo y claro:

- Se delimitan operativamente los objetivos del programa.
- Se recoge información sobre la consecución de dichos objetivos.

- El programa es evaluado positivamente si se han conseguido los objetivos en el nivel, grado y calidad prefijados; si no, se evalúa negativamente.

No siempre es posible utilizar los objetivos del programa como criterio de valor y, en cualquier caso, debe distinguirse entre **crit**erios de valor en sentido amplio y criterios de valor en sentido estricto. En los párrafos anteriores, se tratan los criterios de valor en sentido amplio:

- eficiencia,
- efectividad.. .

Pero también pueden y deben especificarse los criterios de valor más específicos, de existir éstos.

En el ejemplo del sistema de información de la ONCE, la pregunta sobre calidad de la atención plantea qué criterios de calidad utilizar:

- ¿Calidad como satisfacción del usuario?
- ¿Calidad como atención técnicamente correcta?

Preguntas y criterios de valor en la evaluación de MABEM

En la evaluación efectuada para el Instituto de la Mujer sobre el programa MABEM (Método de Aprendizaje de Búsqueda de Empleo para la Mujer), la “cliente” planteó interrogantes/preguntas sobre el programa, centradas en:

- ¿Consigue realmente MABEM facilitar la búsqueda de trabajo a las mujeres?
- ¿Compensa esta efectividad el esfuerzo económico y de gestión que realiza el Insitituto?
- ¿Realmente es el Instituto de la Mujer el que debe encargarse de este programa?

Las preguntas que el Instituto de la Mujer quena responder en la evaluación son preguntas que giran implícitamente en torno a los criterios de valor de eficiencia y **eficacia/efectividad**. Se busca saber si MABEM realmente logra los efectos perseguidos —acercar los puestos de trabajo a las mujeres que los buscan (efectividad)— y si los efectos conseguidos compensan el coste del programa (eficiencia).

La última pregunta es una mezcla de eficiencia y eficacia, puesto que hace referencia al hecho de que el Instituto de la Mujer **podría** dedicar el dinero del MABEM a otros programas y dejar éste en manos de otras instituciones dedicadas a promover el empleo, como el INEM.

Se podrían haber planteado otras preguntas y otros criterios de valor que hubieran marcado otro tipo de evaluación. **Así**, se podría haber planteado la necesidad de difundir la imagen del Instituto y la creación de una imagen de eficacia del mismo, de cara a la opinión pública y las mujeres españolas. Estos nuevos objetivos hubieran podido surgir directamente como tales o como preguntas sobre el programa, y dan'an lugar a explicitar otros criterios **de valor** distintos de los de **eficiencia/eficacia** y otros **tipos** de evaluación.

De nuevo cabe plantearse si el evaluador debe centrarse exclusivamente en las preguntas (y por tanto en los criterios de valor, objetivos y tipos de evaluación) marcadas por el cliente.

Las dos posturas extremas serían:

- Adoptar un enfoque descriptivo, es decir, acatar claramente los "deseos del cliente".
- Adoptar un enfoque prescriptivo y utilizar otros criterios de valor.

El problema que surge de adoptar la segunda postura es la justificación de unos determinados **criterios** de valor, justificación que necesariamente será filosófica, ética, pero no empírica. **Así**, por ejemplo, M. Scriven aboga por utilizar siempre en toda evaluación

la satisfacción de las necesidades de la población objeto o usuarios como criterio de valor, justificándolo desde el utilitarismo y el **igualitarismo**. Resulta evidente que la justificación **pragmática** menos ambigua es la que adopta la primera postura: recurrir a lo que marque el "cliente".

En la práctica, muchas veces el "cliente" no tiene claras ni las preguntas ni los **criterios** de valor, y el evaluador puede ampliar los objetivos de la **evaluación** y sugerir preguntas y/o criterios de valor.

De hecho, tanto el evaluador externo como el interno no tienen otra opción que "adoptar" los criterios de valor de sus clientes (o jefes) con una mayor o menor posibilidad de **negociación**, aclaración o discriminación según los casos.

La disyuntiva entre enfoque "**descriptivo**" y enfoque "**prescriptivo**" se **sitúa** fuera del momento en que se produce un encargo de evaluación; se **sitúa** en la propia metodología de la evaluación. ¿Es posible proponer ciertos **criterios** de valor que necesariamente **deberían** utilizarse o no? O, dicho de otra manera, ¿es posible encontrar justificación a la utilización de determinados criterios de valor y no de otros?

Si la respuesta es sí a ambas preguntas, estamos adoptando una postura **prescriptiva/normativa** al respecto; y esto es lo que hacen muchos profesionales. Por ejemplo, los economistas insisten en que el único criterio de valor es la eficiencia; los interventores españoles suelen utilizar como **criterio** el formalismo normativo.

En las sociedades en que vivimos, se propugna la vigencia de una pluralidad de valores (libertad, justicia, igualdad...), por lo que resulta evidente que debe adoptarse una postura **descriptiva** y **pragmática**, no normativa, al respecto.

2. El diseño **propriadamente** dicho

Un diseño es la lógica subyacente a una investigación —o un **proyecto** de investigación— que relaciona los objetivos de la **investigación**, los datos y su análisis e interpretación.

Cuando se intenta buscar un diseño de evaluación, lo que buscamos es que esa conexión entre datos, análisis y objetivos sea coherente y consistente; o, dicho de otra manera, queremos un plan de recogida de información y de análisis que permita cubrir adecuadamente los objetivos de la evaluación y, por tanto, responder adecuadamente a las preguntas de las que surge la misma.

Un diseño debe incluir:

- las preguntas;
- las hipótesis **y/u** objetivos;
- las unidades de análisis;
- la lógica que conecta los datos con las **hipótesis/objetivos**;
- los criterios para interpretar los datos.

De una manera más concreta, se trata de precisar:

- las preguntas de la evaluación (y objetivos);
- la información necesaria para responder a dichas preguntas (datos) y cómo analizarlas.

Cada tipo de evaluación requiere diseños distintos y, a modo orientativo, se analizan a continuación los diseños más adecuados y frecuentes para distintos tipos de evaluación.

La distinción entre **evaluación interna** y **externa** introduce un criterio pragmático adicional importante en la selección de un diseño específico.

En lo que respecta a los cuatro tipos de evaluaciones mencionadas, cabe señalar que la evaluación del diseño y conceptualización del programa y la evaluación de implementación (de proceso o funcionamiento) suelen realizarse a través de evaluaciones internas. La evaluación de necesidades, la de resultados y la económica frecuentemente se realizan fuera de la organización, por evaluadores externos.

Los criterios clave en la **utilización** de evaluación **interna/externa** son:

- Los conocimientos técnicos exigidos por cada **tipo/diseño** de evaluación.
- Los recursos económicos y organizativos existentes.
- El objetivo de **control/auditoría** que tienen a veces las evaluaciones.

En la evaluación de necesidades, es frecuente la utilización de datos secundarios e, incluso, de técnicas de grupos. Esto es así fundamentalmente cuando se realiza una evaluación interna, por la dificultad de utilizar encuestas a expertos o población en general. Se recurre a evaluadores externos cuando se utilizan estos diseños.

Un análisis similar realizado en **función** de los criterios señalados —recursos, preparación técnica y control— puede hacerse de los demás **diseños** contraponiendo evaluación interna y evaluación externa, pero la decisión sobre llevar a cabo una u otra depende, en **gran medida**, del tipo de organización y de la cultura **evaluadora** existente en **la** misma. En un extremo, hay organizaciones que tienen una amplia práctica y cultura evaluadoras, y prácticamente toda la evaluación es interna, recurriendo muy puntualmente a la evaluación externa. En el otro extremo, existen muchas organizaciones y organismos públicos sin posibilidades técnicas, ni práctica, ni cultura evaluadora, donde cualquier evaluación que se realiza se lleva a cabo externamente.

a. La evaluación de necesidades

Los diseños de investigación utilizados en evaluación de necesidades se centran en el uso de:

- 1) Datos secundarios.
- 2) Encuestas.
- 3) Técnicas de grupos.

El primer punto hace referencia a la utilización de estadísticas provenientes, normalmente, de censos de población pero también de otras **fuentes** institucionales (estadísticas de consumo, de sanidad, de educación...); un caso especial es la utilización de la propia estadística generada por el servicio o programa ya existente.

Este **último** caso sólo es posible en programas o servicios ya existentes cuando se plantea una reforma o modificación de los mismos. El **funcionamiento** de este programa o servicio sirve como base para determinar cuál o **cuáles** son los problemas, y realizar una evaluación de necesidades.

La ventaja clave de utilizar estadísticas (censos o archivos de programas o servicios) es el ahorro de **esfuerzo** y dinero, pero también hay desventajas:

- a) En el caso de estadísticas "oficiales" (censos, padrones), muchas veces el nivel de desagregación no es el adecuado, las variables disponibles no son las más convenientes o, simplemente, son inexistentes y —lo más importante— los datos suelen estar desfasados. En España, el primer avance de resultados del Censo de 1991 aparece en 1994.
- b) En el caso de documentación y archivos de programas/servicios, el problema clave es la ausencia de información sobre no **usuarios/clientes** del **programa/servicio** junto con la necesidad de que la información existente esté completa y en formato adecuado para su utilización.

En cualquier caso, la utilización de datos secundarios tiene ventajas para el logro de determinados objetivos que se especifican en el cuadro siguiente:

| Objetivos de evaluación de necesidades alcanzables según tipo de diseños | |
|---|--|
| Censos/padrones estadísticas oficiales | <ul style="list-style-type: none"> - Análisis de factores de riesgo - Zonificación y distribución de la población de riesgo |
| Archivos/documentos del programa | <ul style="list-style-type: none"> - Análisis de la cobertura y problemas asociados - Tasas de abandono - Accesibilidad del programa/servicio |
| Encuestas | <ul style="list-style-type: none"> - Demanda no cubierta - Necesidades subjetivas - Distribución y volumen de la demanda potencial |
| Técnicas de grupo | <ul style="list-style-type: none"> - Movilización de la población - Avance de soluciones |

Sin duda, la utilización de encuestas es el diseño más flexible, ya que proporciona una mayor información adicional; caben tres variantes de la encuesta según a qué población vaya dirigida:

- a. Encuesta a población general.
- b. Encuesta a expertos.
- c. Encuesta a usuarios del programa/servicio.

En este último caso, sigue existiendo el problema de no disponer de información sobre no usuarios ya señalado, pero en cambio se puede recoger gran cantidad de información de usuarios, mucho más de la que exista en los archivos.

La encuesta a expertos adopta, a veces, algún procedimiento tipo **Delphi**, pero en cualquier caso la idea básica es la utilización de

expertos en el problema que se quiere resolver, que puedan proporcionar información adecuada y previsión sobre el volumen del problema, sus características y sus posibles soluciones. Es evidente que, si estos expertos existen, la encuesta permite recoger información de un modo rápido y eficaz, pero el problema básico de este diseño es la dificultad de identificar a los expertos que, además, muchas veces no existen.

Un ejemplo de utilización del método Delphi en evaluación de necesidades

Durante el año 1989, el Ayuntamiento de Madrid realizó una evaluación de necesidades a la vez que ponía en marcha el Plan Municipal de Lucha contra las Drogas. Los resultados de esta evaluación se encuentran sintetizados en el libro *El consumo de drogas en el Municipio de Madrid*, Ayuntamiento de Madrid, 1990, e incluye un análisis especial de cuatro barrios —Hispanoamérica, Lucero, Concepción y Palmerías— que utiliza el método Delphi.

En cada uno de los cuatro barrios, se identificaron entre 8 y 10 informantes expertos en la situación del consumo de drogas en el barrio (policía municipal, trabajador social, educador de calle, miembros de asociaciones de vecinos, miembros de asociaciones de ex-adictos...). Una vez identificados y contactados, se les aplicó, mediante entrevista personal, un cuestionario totalmente abierto sobre el consumo de drogas en el barrio, su evolución, características de los consumidores y respuesta institucional y ciudadana ante el problema.

El análisis de estas encuestas dio paso a un informe por barrio que luego fue discutido con los expertos en grupos de discusión.

Todo este proceso permitió conocer en profundidad el tema a nivel de estos cuatro barrios, tanto para cuantificar los consumos de las diferentes drogas como para caracterizar al consumidor y la respuesta al problema.

Sin embargo, los resultados obtenidos no cubrieron las expectativas depositadas en el método, porque realmente los supuestos expertos desconocían la situación del barrio. Éste es precisamente el talón de Aquiles del método Delphi, tienen que existir auténticos informantes/expertos y, además, deben poder ser identificados.

La encuesta a población general permite:

- Identificar y caracterizar la población objeto.
- Estimar el volumen de la misma.
- Avanzar algo sobre posibles soluciones.

La encuesta presenta dos posibles problemas que están entrelazados:

- a) No es un buen método de búsqueda de nuevas soluciones; sí puede **servir** para averiguar el apoyo que tiene una solución, pero no es una técnica que genere nuevas ideas, **nuevos temas**, nuevas soluciones.
- b) En la encuesta, el investigador impone su estructura **cognitivo-conceptual** en el cuestionario y, por tanto, en la recogida de información, "produciéndola" en cierta medida. Elaborar buenos cuestionarios es siempre muy difícil y prácticamente imposible cuando la temática no ha sido suficientemente **estructurada** previamente, es decir, cuando se trata de temática novedosa.

La **utilización** de **grupos**, por último, incluye utilización de grupos de discusión, el foro y la asamblea comunitaria. En todos los casos, se trata de diseños que implican participación, por lo que son buenos procedimientos de:

- Motivación y movilización.
- Búsqueda de soluciones consensuadas.

No son métodos muy adecuados para estimar el volumen ni las características de la población objeto del problema y, además, se suelen circunscribir a comunidades reducidas por la imposibilidad de representar en pocos grupos a grandes poblaciones.

Tanto en los grupos como en el foro y la asamblea, los participantes son partes interesadas en la solución de la situación **proble-**

mática y pueden tener además la característica de ser expertos; de aquí que estén capacitados para discutir el problema o situación problemática, **caracterizándola** y avanzando posibles soluciones que sirvan como base de la intervención posterior.

No se busca la representación estadística de estas partes interesadas en los grupos o asambleas, por lo que no cabe hablar de muestras representativas. Tampoco se busca la representatividad a través de estas técnicas de grupo, sino conocer y caracterizar el problema desde la propia población que lo padece. "Sufrir" el problema se constituye en la justificación esencial para dar credibilidad a la información así obtenida, incluidas las posibles soluciones.

Conviene señalar que, aunque se rechace la posibilidad de que la población objeto pueda tener información adecuada y correcta sobre la situación problemática y sus soluciones, es evidente que la información que proporcione matizará la que puedan dar los técnicos y, además, como ya se ha señalado, la participación en la evaluación de las necesidades generará movilización, lo que hará más probable que la intervención tenga éxito.

Los grupos, foro y/o asamblea comunitaria constituyen un complemento de otras perspectivas (la de los técnicos, políticos, etc.) pudiendo:

- Aportar información sobre el problema o sobre posibles soluciones adicionales.
- Servir como movilización de la población.

La utilización de grupos de discusión en la evaluación de necesidades de las mujeres que quieren reincorporarse al trabajo

En 1991, se realizó una evaluación de las necesidades de las mujeres que querían reincorporarse al trabajo en España, para el Instituto de la Mujer. En dicha evaluación, se utilizaron grupos de discusión y encuesta a una muestra representativa de mujeres reentrantes. Los grupos (2 en total) proporcionaron abundante información para caracterizar la problemática de la mujer reentrante y más específicamente:

- Las características básicas de las mujeres reentrantes.
- La imagen que tienen del mercado de trabajo.
- Los factores destacados en el logro del puesto de trabajo.
- Los procedimientos de búsqueda de empleo.

Los grupos sirvieron tanto para aclarar toda la problemática existente como para preparar el cuestionario que con posterioridad se utilizó en la encuesta.

Se eligieron dos grupos de mujeres de edad intermedia (25 a 50 años), inscritas en el paro como representantes prototípicas de la mujer reentrante; no hubo problemas de reclutamiento de los miembros de grupo ni de lograr que elaboraran discursos sobre el tema porque las mujeres se sentían partícipes y motivadas.

b. La evaluación del diseño y la conceptualización de la intervención

Este tipo de evaluación no admite muchas variaciones de planteamiento (diseño), puesto que consiste simplemente en el análisis lógico-conceptual del **programa/intervención** para valorar su adecuación.

Tres son los grandes apartados que debe contemplar una evaluación del diseño/conceptualización:

- a) La coherencia entre el programa diseñado y los resultados de la evaluación de necesidades o, en su ausencia, con las **características** del problema que se quiera solucionar.
- b) La adecuación del programa al modelo teórico de intervención y de éste con los conocimientos teórico-empíricos de las disciplinas básicas sobre el problema.
- c) Una correcta resolución de los problemas típicos de la programación: organización, recursos y previsión para la aplicación de la intervención.

Parece evidente que la primera pregunta que debe realizarse sobre un programa nuevo debe ir dirigida a estimar la adecuación del mismo a las **características** del problema y a todos los resultados obtenidos en la evaluación de necesidades.

Entre otras cosas, habrá que analizar si el programa es coherente con las características de la población (volumen, **distribución** espacial, motivación, dificultades de acceso...) y también si respeta la priorización de necesidades efectuada.

Pero quizá es el segundo apartado no sólo el más difícil sino **también** —y probablemente por ello— el más importante; el análisis del modelo de impacto teórico y su **justificación** en la investigación básica, junto con el análisis de la traducción efectuada del modelo de impacto al modelo de intervención, son los momentos esenciales de este apartado.

El modelo de intervención en las campañas de tráfico

Éste es el problema básico de las campañas informativas/publicitarias realizadas en los medios de comunicación dirigidos, por ejemplo, a reducir el número de accidentes de tráfico o a reducir la gravedad de los mismos.

El modelo de impacto tiene su justificación en los estudios sobre el impacto de los mensajes y la comunicación. Las investigaciones en psicología social ponen de relieve cómo la relación existente entre creencias, actitudes y conducta es, cuando menos, problemática y, en la mayoría de los casos, pequeña. Por tanto, si las investigaciones sobre comunicación subrayan lo difícil que resulta en un mundo de comunicaciones en competencia llamar la atención sobre un mensaje para lograr que la audiencia lo reciba y, si además, sabemos que este proceso está muy débilmente relacionado con el cambio de creencia y actitudes, resulta peregrino esperar un efecto importante de las campañas de información/publicidad debido precisamente a dos hechos:

- La cadena causal es muy larga (atención-comprensión-retención-cambio actitudinal-cambio conductual).
- Las relaciones o efectos en dicha cadena causal son pequeños.

La cuestión reside a veces en que no se conoce **otra** forma de intervención o que es la forma más fácil de hacerlo desde el punto de **vista** político o, incluso, económico, a la vez que se "crea" la apariencia de que se está haciendo algo. Se **trata**, por tanto, a veces, del único modelo de "acción" posible.

En el tercer apartado, entramos en el diseño organizacional del programa en relación con los recursos y objetivos fijados; los diferentes elementos de la organización y la propia organización tienen que ser coherentes con los objetivos que se fijan en la intervención dentro de los **condicionantes** de recursos humanos y financieros.

c. La evaluación de la implementación

Dentro de este tipo de evaluación se recogen, al menos, tres objetivos diferenciados:

- La evaluación de la cobertura.
- La evaluación de la implementación propiamente dicha.
- La evaluación del funcionamiento del **programa/intervención**.

En el primer caso, se trata de analizar si el programa llega a la población objeto, en qué grado y si no hay sesgos en la cobertura; en el segundo, se trata de comparar el **programa/intervención** sobre el papel con el funcionamiento real del mismo y, en el **último** caso, se trata de analizar el propio funcionamiento del programa. **Las** tres dimensiones están muy **interrelacionadas** y son difíciles de separar.

En la evaluación de cobertura, los datos referidos a la población objeto, normalmente obtenidos en la evaluación de necesidades, se comparan con los datos de archivo o del sistema de información del programa, referidos a los usuarios y a sus características; en la evaluación de la implementación o del funcionamiento, debe recogerse información sobre el programa en marcha a través de:

- Observación *in situ* (participante o sistemática).
- Entrevistas o encuestas a usuarios y técnicos o ambos.
- Grupos de discusión con usuarios.

Los problemas de los diseños de este tipo de evaluación se centran en:

- a) La utilización de diferentes fuentes de información para aumentar tanto la validez de constructo como la externa.
- b) Un adecuado diseño muestral de centros — en caso de que el programa tenga más de un centro o servicio — y de tiempos para tener en cuenta los propios desarrollos temporales del programa.

El diseño de la evaluación de la ONCE

El sistema de información para la evaluación de la ONCE es un buen ejemplo de diseño de evaluación de la cobertura y funcionamiento de los programas.

En capítulos anteriores, se han descrito los objetivos/preguntas de la evaluación efectuada desde el sistema de información, que se centran en el análisis de:

- la cobertura de afiliados y características de los usuarios,
- funcionamiento de los diferentes servicios y programas, describiendo actividades, tiempos y actuaciones de los distintos técnicos/profesionales, y análisis de los recursos.

Para cumplir estos objetivos, la dirección de la ONCE decidió elaborar un sistema de información permanente que recogiera la información en los equipos de atención directamente por parte de técnicos y profesionales, que permitiera a la dirección, a través de su gabinete técnico, el análisis de la misma.

La recogida de la información no tenía problemas muestrales puesto que se efectuaba en todos los equipos de atención, pero sí tenía (y tiene) problemas de fiabilidad y validez, ya que es su única fuente que la facilita: los técnicos o profesionales. No hay fuentes externas ni información recogida directamente de los afiliados (usuarios) de los programas.

El problema se agrava además si se piensa que el sistema de información es claramente un sistema de control de la labor de los profesionales y que son los propios profesionales los que tienen que facilitar la información; información que a veces es complicada y costosa de compilar.

Metodológicamente, hubiera sido mejor complementarla con información facilitada por usuarios (afiliados) o evaluadores externos, pero razones de tipo económico y de política interna de la organización obligaron a seguir el otro diseño.

Los posibles problemas de validez de los datos se han intentado paliar mediante tres procedimientos:

1. Análisis interno comparativo de los equipos para detectar "casos extraños".
2. Visitas a los equipos por parte de los técnicos de la dirección general.
3. Reuniones con los encargados directamente responsables de los equipos de atención.

d. La evaluación de resultados y de impacto

En ambos tipos de evaluación, los buenos diseños giran alrededor de la **lógica** experimental puesto que, en ambos, el objetivo esencial es establecer relaciones de causa-efecto **entre** el programa y los resultados; debe primar la validez interna en la elección de un diseño y, por ello, deben **elegirse** diseños experimentales o cuasi experimentales.

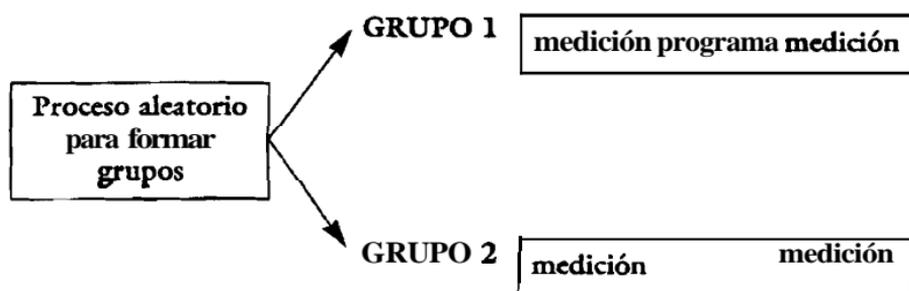
En este sentido, quiero destacar el diseño experimental más clásico y **sencillo** (*pretest-postest* con grupo de control equivalente) y dos diseños cuasi experimentales:

- Series temporales interrumpidas.
- *Pretest-postest* con grupo de comparación no equivalente.

Diseños experimentales

La característica básica de los **diseños** experimentales es la utilización de **grupos** de control equivalentes que, necesariamente, tienen que ser formados a través de **algún** procedimiento aleatorio; el programa constituye la manipulación o tratamiento y el **quid** del problema está en buscar un sistema para lograr esos grupos de control iguales. La medición *pretest* no es estrictamente necesaria pero sí conveniente.

El esquema de este diseño es el siguiente:



Las mediciones antes del programa no son **estrictamente** necesarias, pero sí es necesario el procedimiento aleatorio de formación, de al menos, dos grupos iguales. Puede haber más de un grupo de control y más de un grupo al que se le aplique el programa (experimental). Asimismo, puede haber más de una medición después de la aplicación del programa. Lo esencial en los diseños experimentales es la igualación del grupo de control y del experimental, lo que permite aislar el efecto del programa si en las mediciones que efectuamos encontramos diferencias entre ambos grupos.

Dos grupos iguales en todo, uno de **ellos** recibe el programa y el otro no, e inmediatamente después se recoge información de ambos y se compara para ver si hay diferencias; de haber diferencias, se tienen que deber al programa; si no las hay, el programa no es efectivo.

El diseño de evaluación del programa de autocuidados en asma

Durante los años 1993 y 1994, Amparo Mancebo ha efectuado, en la Consejería de Sanidad de la Comunidad Autónoma de Madrid, una evaluación de un programa de formación en autocuidados para asmáticos utilizando un diseño experimental.

Se quería averiguar si un programa de autocuidados en asma permitía:

- elevar la calidad de vida de los asmáticos;
- disminuir sus visitas al servicio de urgencias y hospitalizaciones;
- mejorar su autoestima y capacidad de hacer frente a la enfermedad.

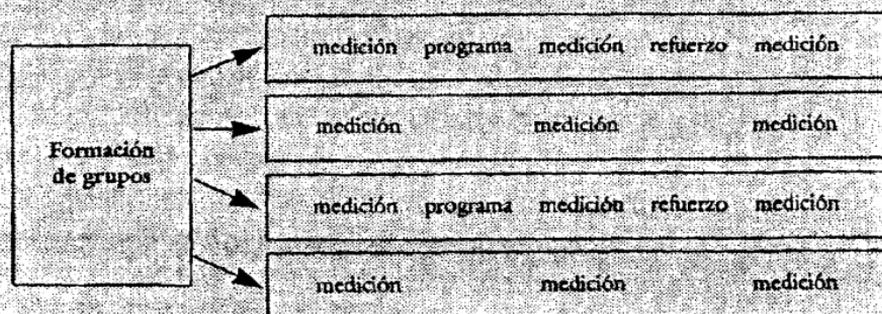
Puesto que el programa se quería implantar en toda la Comunidad de Madrid y era mucho lo que estaba en juego, se pensó en llevar a cabo un experimento que, de ser positivo, se generalizaría.

Para ello, se formaron dos grupos de control y dos experimentales introduciendo las variables:

1) gravedad de la enfermedad, distribuyendo entre severos y leves; y 2) edad de los pacientes.

Los grupos de control y experimentales se formaron aleatoriamente utilizando las fichas de enfermos existentes en diferentes ambulatorios, centros de salud y hospitales; con posterioridad, se pidió la conformidad de los enfermos para participar en el experimento.

Las mediciones, el programa y los refuerzos fueron los siguientes:



De tal modo que, después del programa de formación inicial, hubo dos refuerzos y tres mediciones efectuadas a los cuatro grupos. Las

mediciones se basan en recogida de información a través de las historias clínicas y de cuestionarios ad hoc a los enfermos.

Así, teniendo grupos de control y experimentales, y además con medición repetida, es posible aislar el efecto del programa. Por otra parte, al dividir entre severos y leves, cabe analizar si existen efectos interactivos, es decir, si el programa es igual de eficaz para ambos tipos de enfermos o es más eficaz entre los severos, que es la hipótesis inicial.

Diseños cuasi experimentales

Cuando no es posible lograr grupos de control equivalentes, entonces la medición *pretest* sí que es necesaria para analizar las diferencias entre grupo de tratamiento y grupo de control, y poderlas controlar en el análisis; además, para aumentar la validez interna⁵ de este diseño, es recomendable identificar posibles variables perturbadoras e introducirlas en el diseño para controlarlas en el análisis.

El diseño de evaluación del MABEM

La evaluación del MABEM es un buen ejemplo de diseño cuasi experimental de grupos de comparación no equivalentes.

Los equipos de empleo encargados del MABEM utilizaban un diseño preexperimental⁶ de medición antes y después, con seguimiento a seis meses, sin grupo de comparación, pensando que la repetición de la medición era suficiente control.

| | | | |
|----------|----------|----------|----------|
| medición | programa | medición | medición |
|----------|----------|----------|----------|

⁵ Se entiende por validez interna, siguiendo la perspectiva sistematizada por D. Campbell y colaboradores, el control de las hipótesis alternativas o variables perturbadoras.

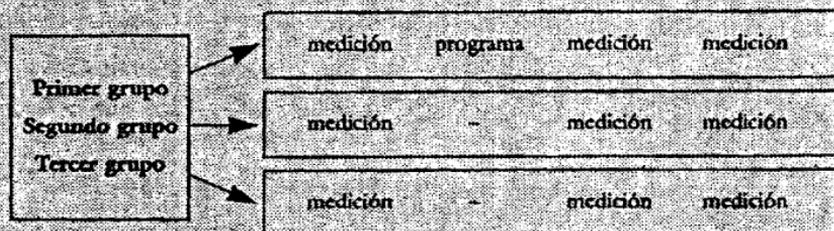
⁶ Un diseño preexperimental es un diseño con muy escaso control de variables perturbadoras, lo que le aleja de los experimentales y de los cuasi experimentales.

Resulta evidente que, con ese tipo de diseño, es imposible interpretar los resultados de logro de trabajo como debidos al MABEM, puesto que hay muchas hipótesis alternativas⁷ no controladas: la pregunta clave es cómo sabemos que las mujeres que siguieron el MABEM no habrían logrado trabajo en la misma proporción de no haberlo seguido. Simplemente, no lo sabemos porque no hay un grupo cero, un grupo de comparación que no siga el MABEM.

Por ello, al comenzar la evaluación, la única vía de aumentar la validez de la misma era buscar grupos de comparación; dado que era una evaluación externa, realizada después de que el programa estuviera funcionando, era imposible buscar grupos equivalentes inicialmente formados mediante un procedimiento aleatorio.

Se logró formar y encontrar información sobre dos grupos de comparación: el primero, una muestra de mujeres que en la fecha del inicio del programa (antes de) estaban registradas en el INEM y tenían la misma edad y residían en las mismas provincias que las mujeres que siguieron el MABEM. El segundo, un pequeño grupo de mujeres que, aun habiendo solicitado participar en el MABEM y habiendo sido seleccionadas, luego no lo realizaron. Este último grupo era casi idéntico al grupo de mujeres que siguieron el MABEM, como mostró luego la comparación de las variables sociodemocráticas de ambos grupos.

Por tanto, el diseño final fue el de la medición *pretest/postest* con dos grupos de comparación no equivalentes.



El diseño de series temporales consiste en la repetición de las mediciones antes del tratamiento para lograr una buena base de comparación y repetición, asimismo, de las mediciones después del tratamiento. Cuantas más mediciones se puedan efectuar, mejor,

⁷ Hipótesis alternativas significa lo mismo que variables perturbadoras.

puesto que esto permite aislar efectos ajenos al tratamiento, al establecer mejor la base de la comparación previa al tratamiento, identificar diferencias introducidas por éste.

El diseño de evaluación de las campañas de tráfico

En 1993, uno de los autores del libro⁸ realizó un estudio para evaluar el efecto de las campañas de información en los medios de comunicación, patrocinadas por la Dirección General de Tráfico con el objetivo de reducir:

- El número de accidentes de tráfico.
- La gravedad de los mismos.

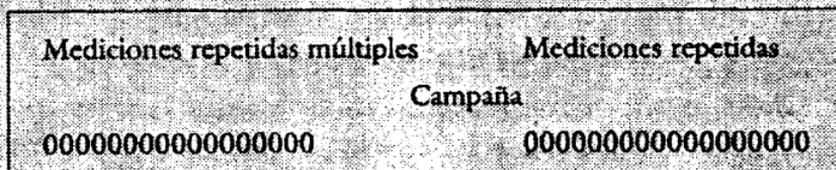
En este caso, un diseño auténticamente experimental hubiera sido, metodológicamente hablando, lo mejor pero desde el punto de vista pragmático éste era imposible.

Por ello, se eligió utilizar un diseño de series temporales interrumpidas, a la vista de que la DGT disponía de la información sobre accidentes día a día en un período superior a tres años; es decir, a la vista de que se disponía de más de 50 puntos antes de la intervención y de 50 puntos después.

Cabía haber buscado un grupo de comparación no equivalente, grupo que no viera las campañas de publicidad pero esto resultaba costoso y no era estrictamente necesario.

Para cada accidente se disponía de más de 80 variables/características del mismo: edad del conductor, sexo, características del vehículo, de la carretera, del día..., por lo que el propio análisis podría servir para controlar hipótesis alternativas y aislar el posible efecto de las campañas.

El diseño se esquematiza de la siguiente manera:



⁸ F. Alvira y F. Blanco hicieron la evaluación a que se alude aquí; a partir de esta fecha, F. Blanco trabaja con datos de tráfico en su tesis sobre los modelos ARIMA y la estimación de efectos.

Los dos procedimientos utilizados para aumentar la validez interna del diseño fueron:

- Modelización de la serie antes de la campaña e introducción de la campaña como variable independiente después.
- Utilización de variables dependientes diferentes que maximizarían la comparación, por ejemplo, separando las series de accidentes mortales, accidentes leves, accidentes de coches, de motos..., con lo que realmente se consiguen casi series de comparación no equivalentes.

e. La evaluación económica

Los diseños utilizados en evaluación económica giran en torno al problema básico de este tipo de evaluación: la transformación de *inputs* (tratamiento o programa) y de *outputs* (resultados, efectos) a términos monetarios. El énfasis se pone en la correcta identificación de todos los costes y también en la correcta valoración de los mismos términos de dinero; el mismo énfasis se pone en la identificación y valoración de los efectos o resultados. Dado que este tipo de evaluación requiere unas técnicas económicas muy específicas, en este libro no se trata este tipo de evaluación.¹ Sólo se mencionan los diferentes tipos de evaluación económica.

La evaluación o análisis **coste-beneficio** es, sin duda, el tipo de evaluación económica tradicional en el que los costes del programa o intervención se comparan con el beneficio (en términos monetarios) que provoca la intervención; esto supone normalmente comparar varias alternativas de intervención o programas.

Cuando transformar a términos monetarios los efectos de una intervención resulta **difícil** o incluso imposible, se utiliza el análisis **coste efectividad**; en el lado de los efectos, simplemente se ponen

¹ Para un desarrollo mayor de estos métodos de evaluación económica, puede consultarse M. J. Aguilar y E. Ander-Egg, *Evaluación de servicios y programas sociales*, Buenos Aires, Lumen, 1994 (N. del E.)



las unidades de producto. Por ejemplo, un análisis económico de diferentes programas de rehabilitación de drogadictos se puede hacer comparando los costes por drogadicto rehabilitado. El efecto es la rehabilitación y el producto es "un drogadicto rehabilitado"; esto permite comparar los programas de rehabilitación en términos de eficiencia, es decir, en términos de costes o buena utilización de los recursos.

Por **último**, cabe hablar del **análisis** coste-utilidad, en el que los efectos se transforman a términos de utilidad subjetiva. Es un análisis intermedio entre el análisis coste-beneficio y el coste-efectividad, y requiere un procedimiento para asignar **utilidades**.