

Anexo 6. Técnicas de investigación

Este anexo pretende aportar información sobre diferentes metodologías que pueden ser de utilidad a los equipos implementadores. Atendiendo a su naturaleza, algunas de ellas se aproximan de forma explícita a la implementación de GPC, mientras que otras se abordan de forma más genérica ofreciendo una visión general, dada su posible utilidad y diversos objetivos en diferentes momentos del plan de implementación.

No se pretende proponer un listado exhaustivo de técnicas de investigación de diferentes metodologías ni profundizar en su conocimiento; tan sólo señalar aspectos que son necesarios conocer, o sobre los que se debe reflexionar, a la hora de poner en marcha procesos de implementación. Por supuesto, a la hora de utilizar cualquier técnica de investigación es necesario recurrir a profesionales expertos en su manejo.

1. Técnicas de consenso

En diversos momentos de la elaboración o la implementación de una GPC se requiere que los miembros del grupo elaborador tomen decisiones colectivas, basadas en acuerdos comunes. Para alcanzar este tipo de acuerdos y decisiones existen diferentes técnicas de consenso que pueden ayudar a hacerlo de forma organizada y sistemática.

Consensuar, según la Real Academia de la Lengua, significa “adoptar una decisión de común acuerdo entre dos o más partes”.

Para garantizar la validez de un consenso resulta fundamental:

- Establecer cuestiones a responder y objetivos concretos y específicos.
- Seleccionar el grupo de expertos de manera que se garanticen aspectos como un número suficiente de miembros, experiencia, prestigio, interés por el tema, disponibilidad de tiempo e inexistencia de conflictos de interés.
- Los escenarios deben ser elaborados metódicamente, siguiendo un proceso formal estructurado.

Sin embargo, incluso dentro de grupos homogéneos, puede ser complejo alcanzar acuerdos o decisiones, debido a la diferencia de conocimientos de los miembros, la fuerza de las pruebas aportadas o incluso los intereses particulares.

Es importante definir qué se considera “acuerdo” del grupo y establecer la forma de medir el grado de acuerdo.

Las principales características de los métodos de consenso son el anonimato, la iteración, la *retroalimentación* controlada y la conclusión de una respuesta concreta y resumida.

Para elegir entre los diferentes métodos participativos se deben considerar los siguientes aspectos:

- Objetivos: razones, aspiraciones, motivaciones y el fin último que se quiere lograr.
- Conocimiento, complejidad, controversia y contexto del tema a tratar.
- Participantes que se pueden reclutar.
- Tiempo disponible.
- Presupuesto.
- Ventajas e inconvenientes de cada técnica.

1.1. Técnica de Delphi

No es necesaria la presencia física de los expertos, con los que se puede contactar por correo y sin existir criterios estandarizados sobre el número idóneo de participantes, dependiendo este aspecto de la naturaleza del grupo, de los objetivos, etc.

Se realizan varios envíos de un cuestionario a los participantes, habitualmente llamadas rondas, preguntándoles, en la primera de ellas, su opinión acerca de un tema determinado. En envíos sucesivos los participantes reciben un resumen de los resultados de las rondas anteriores con la posibilidad de revisar sus propias opiniones. Las respuestas de los participantes se agregan mediante métodos estadísticos.

Esta técnica implica abordar las siguientes fases:

1. Definición del problema sobre el que alcanzar consenso.
2. Selección de los expertos.
3. Diseño de la prueba preliminar del cuestionario, incluyendo una prueba piloto que evalúe la ambigüedad.
4. Desarrollo de la primera ronda: los expertos cumplimentan las preguntas del cuestionario y la reenvían a los encargados de su análisis.
5. Análisis de las respuestas de la primera ronda (mediana y rangos intercuartílicos para cada pregunta) que son enviados a los expertos.
6. Preparación y envío de los cuestionarios de la segunda ronda, a fin de darles a conocer la opinión global para cada pregunta y disminuir la variabilidad de las opiniones, reasignando una nueva calificación para aquellas preguntas que quedaron fuera del rango.
7. Análisis de las respuestas de las siguientes rondas, que son enviadas de nuevo a los participantes con el fin de establecer el grado de consenso.
8. Preparación del informe final y distribución entre los participantes.

El éxito de un Delphi dependerá del número de rondas y el tipo de *retroalimentación*.

Al no ser imprescindible la presencia física de los expertos, se eliminan los problemas derivados de la distancia y las complejidades de organizar reuniones presenciales, y se garantiza la confidencialidad, reduciendo las posibles influencias entre los miembros. Además, la fluidez del intercambio de información enriquece las conclusiones definitivas. No existe una guía universal de aplicación y constituye un proceso laborioso.

1.2. Técnica de grupo nominal

Técnica de priorización mediante la cual se obtienen resultados de una forma rápida, aunque alcanzando un consenso menor que mediante la técnica Delphi.

En rondas de trabajo, el grupo de expertos, compuesto por 9-12 personas, discute en presencia de un coordinador. En una primera fase, los expertos elaboran preguntas sobre un tema de manera individual. Posteriormente, los expertos se reúnen para discutir acerca de las cuestiones que ellos plantearon. Finalmente, cada experto, de forma confidencial e individual, valora por escrito las propuestas que le parecen más interesantes.

También es posible responder y priorizar las preguntas propuestas por el moderador. Tras la discusión, cada experto puntúa las respuestas según su importancia. Las puntuaciones que estén fuera del rango intercuartílico son discutidas con el grupo, que tiene la posibilidad de cambiar su puntuación inicial.

La gran interacción entre los expertos permite profundizar en algunas cuestiones y su aplicación generalmente en un único objetivo, son parte de las ventajas del método, así como el hecho de que su desarrollo resulta más rápido que el Delphi. El carácter presencial de esta técnica puede dar lugar a ciertos sesgos asociados a la presencia de personas con mayor conocimiento o liderazgo en la discusión.

1.3. Conferencia de consenso

Un grupo organizador será el encargado de nombrar un grupo formado por unos 20 expertos en el tema a tratar y un jurado multidisciplinar de 8-16 profesionales, así como de elaborar las preguntas a tratar sobre el tema. El jurado elaborará el texto final, donde quedará reflejado el grado de acuerdo alcanzado, así como los puntos de acuerdo y desacuerdo.

Tras la interacción entre expertos, donde cada uno de ellos puede expresar su opinión durante 15 minutos aproximadamente, el jurado elaborará un texto con las conclusiones, a puerta cerrada, de forma independiente y lo más objetiva posible.

La conferencia de consenso resulta útil para la toma de decisiones de difícil abordaje de forma eficiente, a pesar de que para su desarrollo completo se requiere un largo periodo de tiempo. Al igual que en la técnica de grupo nominal, la presencia de personas con mayores conocimientos o mayor liderazgo puede influir en el resultado final, con la diferencia de que aquí los expertos no toman parte de las decisiones finales.

Resulta necesario ser muy preciso en el seguimiento de los métodos y procesos a fin de garantizar la imparcialidad de las conclusiones, ya que este método es poco apropiado para sintetizar el juicio individual de cada participante y está muy influenciado por factores externos y conflictos de intereses.

1.4. Tormenta de ideas (*brainstorming*)

Lo ideal es que el grupo lo formen menos de 12 participantes. Su objetivo es la generación de nuevas ideas a partir de la propuesta de un moderador, cuya misión es estimular la creatividad del grupo. Se deben evitar las valoraciones personales durante la sesión. Es de utilidad para conocer las expectativas de los participantes ante algún evento.

2. Técnicas de investigación cualitativa

El objetivo principal de las técnicas de investigación cualitativa no es demostrar empíricamente los hechos, sino analizar e interpretar el sentido, las motivaciones y los comportamientos vinculados a tales hechos. Para ello es preciso estudiar, con una visión holística, las asociaciones y relaciones que se producen entre diferentes contextos, bien sean estructurales o situacionales, su sistema de relaciones y su estructura dinámica. En definitiva, la investigación cualitativa pretende responder a preguntas relacionadas con el cómo o el porqué de los hechos.

La investigación cualitativa se caracteriza por lo siguiente:

- Enfatizar la importancia del contexto y la valoración holística del fenómeno a estudiar.
- Situarse en el lugar de los individuos para adquirir su perspectiva.

- Describir los escenarios del estudio.
- Enfatizar los procesos y la dinámica natural de la vida social.
- Mostrar flexibilidad en el diseño y renunciar a la utilización de marcos e hipótesis de partida cerradas.

La aplicabilidad de técnicas de investigación cualitativa en el ámbito de las GPC va desde la elaboración (por ejemplo, en la identificación de posiciones de pacientes ante determinados problemas de salud) hasta las diferentes fases de los procesos de implementación (por ejemplo, en la identificación de barreras y facilitadores).

2.1. El muestreo en la investigación cualitativa

El muestreo en la investigación cualitativa no responde a criterios probabilísticos y de aleatorización, sino a criterios teóricos o intencionales, y tiene como objetivo alcanzar el mayor grado de información posible sobre el objeto de la investigación. Por tanto, la unidad de muestreo no son los individuos en sí mismos, sino los conceptos o cuerpos teóricos relacionados con la investigación. En este sentido, el objeto del muestreo no es garantizar la representatividad y generalización de los resultados, sino descubrir significados o realidades múltiples.

El tamaño muestral será flexible, asumiendo que el proceso es acumulativo y que el objetivo final es alcanzar la saturación y, por tanto, estará condicionado por la propia dinámica de la investigación.

Algunos de los tipos de muestreos más utilizados son los siguientes:

- Muestreo de casos extremos: Se basa en la búsqueda de casos que debido a sus circunstancias inusuales o especiales son ricos en información.
- Muestreo de variación máxima: Utilizado cuando las realidades humanas sobre las que pivota la investigación son estables, se centra en la búsqueda de “casos tipo” habituales en el contexto investigado para identificar principales características de dicho contexto.
- Muestreo de casos homogéneos: Se caracteriza por la identificación de sujetos con experiencias comunes sobre el tema de investigación, con el objeto de realizar descripciones profundas de los subgrupos que lo componen.
- Muestreo de caso típico: Fundamentado en la identificación de casos habituales con el objeto de describir rasgos comunes de la realidad objeto de análisis.
- Muestreo de caso crítico: Basado en el planteamiento de una situación, fenómeno o realidad hipotética, con el objetivo de acercarnos a su comprensión.
- Muestreo en cadena o bola de nieve: Un sujeto perteneciente al colectivo objeto de la investigación facilita el contacto con el siguiente, este con el próximo, etc.
- Muestreo por criterio lógico: Fundamentado en la necesidad de contar en la investigación con sujetos que reúnan determinados criterios considerados esenciales en el objeto de la investigación.
- Muestreo por conveniencia: En determinadas ocasiones, circunstancias concretas del investigador y/o de los sujetos investigados pueden requerir obtener la mejor información en el menor tiempo posible, por lo que es necesario optimizar al máximo los recursos disponibles.

2.2. Descripción de las técnicas de investigación cualitativa

Existen diferentes tipos de diseños y técnicas de investigación cualitativa. A continuación se describen brevemente las más utilizadas y con mayor aplicabilidad en el ámbito de la salud, que pueden ser agrupadas en técnicas conversacionales, observacionales y documentales.

Técnicas conversacionales

Entrevista individual

- Estructurada: Consiste en formular de manera estructurada, del mismo modo y en el mismo orden, una serie de preguntas preparadas de antemano a los participantes en la investigación.
- Semiestructurada: Las preguntas son definidas por el investigador a modo de guión y su formulación puede variar en función del desarrollo de la entrevista; puede profundizarse en cuestiones relevantes mediante la formulación de nuevas preguntas.
- En profundidad: Combina la formulación de preguntas abiertas y específicas; prevalece el carácter abierto de la entrevista, basándose en la orientación aportada por el propio entrevistado. La sutil dirección del investigador debe centrarse en formular aquellas preguntas que surjan de las propias respuestas del entrevistado. Esta entrevista permite abordar un número limitado de temas en mayor profundidad que las estructuradas y semiestructuradas.

Técnicas grupales

- Grupos de discusión: Reunión de grupos homogéneos de entre 6 y 10 personas seleccionadas de acuerdo a los criterios de muestreo de la investigación. Mediante la discusión grupal de la temática a abordar es posible conocer y “reconstruir” los valores, las representaciones culturales, normas de referencia, aspectos motivacionales, etc. dominantes en un determinado grupo o estrato social. La participación del moderador del grupo de discusión debe ser limitada y orientada a favorecer un espacio en el que hablar con comodidad y libertad, facilitando la fluidez y circulación de discursos con el objeto final de alcanzar un discurso grupal en el que se recojan las interacciones e intersubjetividad del grupo.
- Grupos focales: Aunque con frecuencia se utiliza indistintamente los términos “grupo de discusión” y “grupo focal”, es necesario apuntar que no son exactamente equiparables. El grupo focal se centra más en el individuo y en su discurso, pudiendo entenderse como una entrevista grupal en la que se escucha en grupo, pero se habla como individuo.

Técnicas observacionales

- Observación participante: Se fundamenta en la integración del investigador (observador) en el espacio y en la vida cotidiana del objeto de la investigación, con el objetivo de profundizar en el conocimiento del mismo. El investigador participa de forma activa en la cotidianeidad del grupo, asumiendo uno o más roles, e identificándose con él hasta el punto de que el grupo lo considere como uno más

de sus miembros. La inmersión en la realidad o hecho a investigar permite un alto grado de comprensión empática del mismo. Esta técnica es especialmente útil para conocer el funcionamiento y las relaciones que se producen dentro de organizaciones o grupos sociales.

- Observación no participante: El investigador (observador) se limita a recoger información del objeto de la investigación sin interactuar con él, sin participar de forma activa dentro del grupo que observa. El observador entra en contacto con la realidad o hecho a estudiar permaneciendo ajeno a ella, adquiriendo el papel de mero espectador.

Técnicas documentales

Consisten en la identificación, recogida y análisis de documentos relacionados con el hecho o contexto estudiado. Estas técnicas pueden constituir el punto de entrada al campo a investigar, pudiendo incluso dar pie a plantear el objeto de investigación en sí mismo, o a modificarlo.

Los documentos de consulta pueden ser de diferente índole (personales, oficiales, informales, grupales, etc.) y abarcan cualquier tipo de documento (revistas, informes, fotografías, etc.).

3. Técnicas de investigación cuantitativa

El evaluador de la implementación de una guía estará interesado en evaluar si las actividades de implementación se han realizado, si los indicadores de proceso se han cumplido, o si la implementación de la guía ha mejorado la salud de los pacientes.

Para el primer y el segundo tipo de indicadores, estimar una simple media o un porcentaje será suficiente; a lo sumo sería útil estimar el intervalo de confianza. En cambio, si lo que pretendemos averiguar es **si la implementación de la guía es la causa de las mejoras de salud de los pacientes**, será preciso utilizar un arsenal analítico diferente.

En este caso, el evaluador deja de actuar en clave de evaluación administrativa y pasa a realizar investigación evaluativa, es decir, a contrastar la hipótesis de que los pacientes han mejorado por efecto de la implementación de la guía.

Los ingredientes básicos de la investigación evaluativa son el diseño de investigación, las hipótesis, el tamaño muestral, las medidas de asociación y el control de las explicaciones alternativas. Todos ellos están íntimamente relacionados, aunque para los objetivos de este anexo se abordarán exclusivamente diferentes aspectos relacionados con el tamaño muestral y las medidas de asociación.

La evidencia experimental o cuasi experimental, la que en definitiva provee la investigación evaluativa, es uno más de los criterios a considerar cuando se quiere atribuir a la implementación de una guía las mejoras de salud obtenidas en los pacientes. Otros criterios, como la alta validez interna del diseño, la coherencia con los resultados obtenidos en experiencias semejantes, la consistencia de los resultados en subgrupos de pacientes o situaciones diferentes, el efecto dosis-respuesta, y la ausencia de explicaciones alternativas deberán ser igualmente considerados para atribuir causalidad a la implementación de la guía.

3.1. El tamaño muestral

El tamaño muestral en la investigación evaluativa está indisolublemente relacionado con la hipótesis de trabajo que el evaluador maneja. Siempre que el evaluador esté interesado en responder una pregunta que contenga expresiones del tipo “más que”, “mejor que”, “atribuible a”, “causado por”, etc. debe formularse una hipótesis de trabajo que permita estimar **el tamaño muestral necesario para encontrar el efecto esperado si este existe**.

Aunque hay distintas formas de estimar el tamaño muestral necesario, según el tipo de medida de efecto, el proceso tiene varios pasos comunes:

- **Describir la hipótesis nula (y su alternativa).** En general, es la hipótesis de la no diferencia entre los grupos de comparación. En el caso de un solo grupo, la no diferencia entre dos medidas realizadas en distinto momento. La hipótesis alternativa es la que establece que la intervención es la que provoca las diferencias entre los grupos, o en el grupo antes y después de la misma. En el tema que nos ocupa, la hipótesis alternativa se formularía del siguiente modo: “La implementación mejoró los resultados de los pacientes”.
- **Determinar el tipo de variable para el resultado principal y para la variable explicativa fundamental.** El arsenal estadístico que está debajo del cálculo del tamaño muestral depende del tipo de variables que se van a utilizar. En la siguiente tabla se presenta una generalización de esta idea.

Tabla A6.1. Estadísticos utilizados en el cálculo del tamaño muestral

		Variable resultado	
		Dicotómica	Continua
Variable predictora	Dicotómica	Ji cuadrado	T test
	Continua	T test	Coef. correlación

En el caso que nos ocupa, la variable predictora fundamental será la implementación o no de la guía (dicotómica); mientras que la variable resultados podrá variar (media de hemoglobina glicosilada, proporción de pacientes tratados, etc.).

- **Determinar la magnitud del efecto esperado y su variabilidad.** Es el aspecto más delicado para la estimación del tamaño muestral. La probabilidad de que un determinado estudio pueda detectar la asociación entre un predictor y un resultado depende de la magnitud con que esta asociación se da en la realidad. Si el efecto es grande, será fácilmente detectable (poco tamaño muestral preciso), mientras que si el efecto es pequeño será difícilmente detectable (exigirá mayor tamaño muestral). Habitualmente, el investigador no sabe cómo de grande o pequeño es el efecto esperado; de hecho, ese es uno de los propósitos de su estudio. El evaluador deberá decidir *a priori* cuál es el efecto que él esperaría de la implementación de la GPC, y qué resultaría clínicamente relevante. Para ello, en general, puede utilizar tres estrategias: aplicar el conocimiento de estudios previos similares, realizar una prueba piloto previa o determinar qué podría considerarse la “mínima mejora clínica” atribuible a la implementación.

Cuando se miden distintos resultados derivados de la implementación (pacientes que mejoran su hemoglobina glicosilada, mujeres que reducen su nivel de colesterol, etc.), se pueden seguir dos estrategias para el cálculo del tamaño: a) definir cuál es el resultado más relevante desde el punto de vista clínico, asumiendo el riesgo de no disponer de tamaño muestral suficiente para algunos de los resultados, o b) utilizar el tamaño de efecto esperado que exige mayor tamaño muestral.

- **Seleccionar los errores alfa y beta asumidos en el estudio.** Se trata de una decisión arbitraria, aunque habitualmente se asume un riesgo sobre el error alfa máximo del 5% y un riesgo de error beta máximo del 20%. En realidad, la decisión debería depender del riesgo que el evaluador quiere asumir para cada uno de los riesgos.

En el caso de la implementación de una guía, si el resultado obtenido señala que la guía ha producido mejoras en los pacientes cuando en realidad no lo ha hecho (error tipo I), puede tener como consecuencia continuar con su implementación, aunque no parece servir de nada. ¿Deberíamos reducir la posibilidad de error tipo I a menos del 5%?

En el caso del error tipo II, el resultado de nuestra investigación señalaría que la GPC no ha mejorado a los pacientes cuando en realidad sí lo habría hecho. La consecuencia sería no recomendar el uso de la GPC. ¿Deberíamos reducir la probabilidad de error tipo II a menos del 20%?

- **Situaciones especiales a considerar en el cálculo del tamaño muestral.** En la evaluación de la implementación de guías deben considerarse algunas situaciones características.
 - ***Cuando se estiman diferencias de medias.*** Cuando se utiliza una diferencia de medias, en la estimación del tamaño muestral necesario es tan importante el efecto (diferencia de las medias entre grupos) como la variación de la diferencia. En estos casos, en la estimación del tamaño muestral será preciso utilizar lo que se llama *tamaño de efecto estandarizado*, que se calcula como el cociente entre la diferencia de medias y la desviación estándar de la misma.
 - ***El efecto de casos apareados.*** Con frecuencia la variable resultado es la diferencia en la medida de un mismo paciente o grupo de pacientes, antes y después de la intervención. En este caso, la varianza esperada (dentro del paciente o del grupo) se espera menor, por lo que el tamaño del efecto necesario para estimar el tamaño muestral será menor.
 - ***El efecto de cluster.*** Con frecuencia, los pacientes son muestreados por grupos. De hecho, debemos esperar que los pacientes de un médico tengan una fuerte correlación en cuanto a los resultados de la implementación de una GPC. Este fenómeno, correlación intraclase o *cluster*, exige un tamaño muestral distinto al caso de que los pacientes fuesen considerados observaciones independientes, siendo necesario considerarlo junto con el tamaño de los efectos y los errores alfa y beta asumidos.
 - ***Otras situaciones que pueden requerir aproximación distinta.*** En estudios con variables categóricas, en estudios en lo que se mide es el efecto dosis-respuesta, en estudios que implican seguimiento en el tiempo o en estudios

en los que se van a utilizar técnicas de análisis multivariante, el cálculo de tamaño muestral puede utilizar la aproximación analítica propuesta en la tabla A6.1. En general, para el tipo de problema que analiza el evaluador de la implementación de una GPC, debería ser suficiente. No obstante, existen métodos específicos que aproximan mejor el tamaño muestral preciso para encontrar efecto si éste realmente existe.

3.2. Las medidas de asociación

Para estimar el efecto de una determinada intervención (implementación de una GPC) dispondremos de distintas medidas de asociación (diferencias de medias, de proporciones, de riesgos, o medidas relativas), según el tipo de variable analizada.

Tres precauciones resultarán útiles para manejar apropiadamente las medidas de asociación:

- **Las medidas absolutas son más útiles que las relativas para tener una idea del impacto real de la intervención (implementación de la guía) en la salud de los pacientes.** Imaginemos que el 10% de los pacientes que han sido objeto de la GPC han mejorado sus cifras de presión arterial mientras que sólo el 1% de los no sometidos a la GPC han mejorado. La protección relativa debida a estar expuesto a la guía es 10 veces mayor ($PR=10\%/1\%$). Imaginemos ahora que el porcentaje de pacientes que mejoran en el grupo que recibe la guía es 1%, y el 0,1% en los que no la reciben. La protección relativa vuelve a ser 10 veces más en los pacientes sometidos a la guía. Sin embargo, aunque la protección relativa es la misma (10), en el primer caso, 9 de cada cien pacientes pueden atribuir su mejora al efecto de la guía; en el segundo caso, este efecto se reduce a 9 de cada 1000.
- **Los estimadores centrales (diferencia de medias, riesgo relativo, etc.) son útiles, pero insuficientes para aportar información relevante al evaluador.** De hecho, el estimador central no aporta información sobre si el azar puede explicar la asociación y tampoco sobre la precisión de la medida obtenida. Para obtener ambas se deben estimar los intervalos de confianza de cada estadístico. Por ejemplo, la reducción del porcentaje de factores de riesgo atribuibles a la implementación de una GPC fue 20% (Intervalo de confianza 95% [IC 95%: entre 14% y 27%]; en un segundo caso, la reducción fue la misma 20% (IC 95%: entre 2% y 42%), y en un tercero, fue idéntica 20% (entre -1% y 45%). En todos los casos, la reducción atribuible fue la misma, pero sólo en el primer caso la medida fue precisa (útil clínicamente). En el segundo caso, tenía igual posibilidad de ser 20% que 2% o que 42% (la imprecisión resultó muy poco informativa para tomar una decisión), y en el tercero el intervalo de confianza incluía el valor 0, valor nulo en las medidas absolutas, con lo que el azar no podía descartarse como explicación de las mejoras observadas tras la implementación de la GPC (en las medidas relativas el valor nulo es el 1). En la tabla A6.2 se ofrece un listado de las medidas posibles y sus intervalos de confianza.

Tabla A6.2. Medidas de asociación y su intervalo de confianza

	Estadístico	Estimador central	Intervalo de confianza
Medidas absolutas	Diferencia de medias		
	Diferencia de proporciones		
	Diferencia de riesgos		
Medidas relativas	Riesgo Relativo		
	Odds Ratio		
	Razón de prevalencias		

- El evaluador de la implementación de una GPC habitualmente compara proveedores sanitarios, o el mismo proveedor dos veces en el tiempo. **La comparación entre proveedores exige siempre utilizar técnicas de ajuste de riesgo**, de forma que las diferencias tras la implementación sean atribuibles a la misma y no a la diferencia en los individuos afectados por la misma. Existen distintas aproximaciones técnicas para reducir este potencial problema. En primer lugar, utilizar un diseño aleatorio; en segundo lugar, aparear los casos de uno y otro grupo, y en tercer lugar, analíticamente, mediante el uso de análisis multivariado. Sería preferible realizar la **estimación ajustada de los estadísticos** descritos en la tabla A6.2 **utilizando técnicas multivariantes** (regresión lineal, logística, de riesgos proporcionales, etc. según el tipo de variable respuesta). Como este tipo de estudios tendrán efecto *cluster*, el tipo de técnica preferible será el **análisis multinivel**.

4. Evaluación económica

La evaluación económica (EE) se define como el conjunto de técnicas empleadas para el análisis sistemático de los costes y beneficios de las alternativas comparadas en una toma de decisión concreta. La estrategia básica de la EE es analizar comparativamente la relación entre el coste de los recursos empleados (procedimiento quirúrgico, fármacos, etc.) y los resultados (beneficios) obtenidos de su aplicación, y constituye un pilar fundamental para apoyar la toma de decisiones al respecto.

4.1. Conceptos de eficacia, efectividad y eficiencia

Los conceptos de eficacia, efectividad y eficiencia son fundamentales en el campo de la evaluación de tecnologías, por lo que es conveniente definirlos con claridad.

- La **eficacia** mide la probabilidad de que una persona, en una población definida, se beneficie de la aplicación de una tecnología sanitaria a la resolución de un

problema concreto de salud en condiciones ideales de aplicación. Se establece de forma experimental y su validez es universal.

- La **efectividad** mide la probabilidad de que una persona, en una población definida, se beneficie de la aplicación de una tecnología sanitaria a la resolución de un problema concreto de salud en condiciones reales de aplicación. Las medidas de efectividad no son universales y su generalización dependerá de la población a la que se aplica y de las condiciones en las que se aplica. La diferencia entre eficacia y efectividad dependerá directamente de la diferencia entre condiciones ideales y reales. Por ejemplo, en el tratamiento antituberculoso con poliquimioterapia, 95 de cada 100 pacientes que lo siguen correctamente (condiciones ideales) se curan. Pero ¿qué ocurre si en una población sólo se diagnostican el 10% de los casos nuevos, se trata correctamente al 80% de los diagnosticados y de estos, sólo el 50% cumple correctamente el tratamiento? La efectividad (condiciones reales) de un programa de lucha contra la tuberculosis quedará limitada a un 3,8% ($0,95 \times 0,10 \times 0,80 \times 0,50 = 0,038$).
- El concepto de **eficiencia** incluye, además, los aspectos económicos asociados a la intervención. Se define como la relación existente entre los beneficios obtenidos de la aplicación de una tecnología y los recursos sacrificados para obtenerlos. Al ser un término relativo ha de ser utilizado en comparación con otro: nada es eficiente en sí mismo si no es en términos de comparación (un determinado tratamiento será más eficiente que otro si, a igual coste, obtiene un mayor beneficio, o, a igual beneficio, su coste es menor).

4.2. Metodología de estudio

4.2.1. Objetivo, punto de vista del análisis y descripción de alternativas

La pregunta de evaluación debe especificar el objeto del estudio evaluativo, el punto de vista desde el que se realizará la evaluación y las alternativas que se van a considerar. Por ejemplo, para formular correctamente una pregunta relacionada con el coste de la hospitalización a domicilio habría que considerar los siguientes puntos de vista: a) el hospital, y b) los pacientes y las familias que soportan los gastos. La pregunta formulada del siguiente modo: *¿Es preferible un programa de hospitalización a domicilio, o la atención ya existente en el hospital?* tiene las ventajas de señalar las alternativas a considerar y especificar los posibles puntos de vista desde los cuales se considerarán los posibles costes y beneficios, permitiendo la definición de los objetivos del análisis y evitando que un simple traslado de costes entre las partes sea asimilado a una disminución de los mismos.

4.2.2. Tipo de análisis

La identificación de los costes y su cuantificación en unidades monetarias se realiza de forma similar en todos los tipos de EE. Sin embargo, la naturaleza de los resultados puede diferir ostensiblemente y aconsejar tanto formas diferentes de medición de los mismos, como de su relación con los costes.

Supongamos que deseamos comparar dos programas para la realización de intervenciones quirúrgicas de cataratas en adultos, en el primer caso bajo ingreso hospitalario

convencional y, en el segundo, mediante cirugía sin ingreso. Se ha comprobado que no existen diferencias significativas entre ambos tipos de intervenciones, aunque en el primer caso el paciente requiere estar dos días en el hospital. Nos hallamos ante dos alternativas de idénticos resultados, pero diferentes costes. La EE es esencialmente, en esta situación, la identificación de la alternativa menos costosa. Por ello, este tipo de análisis se denomina ***análisis de minimización de costes***.

Supongamos ahora que nuestro interés se centra en la comparación de la supervivencia en los pacientes con leucemia mieloide aguda, comparando los costes y los resultados del tratamiento quimioterápico convencional frente al trasplante autólogo de médula ósea. En este caso, el resultado (supervivencia) es común a ambos programas; pero, y a diferencia del caso anterior, no solo los costes son distintos, sino que también lo es la efectividad de ambos programas, ya que la prolongación de la vida obtenible es diferente según el tratamiento elegido. En esta situación no podemos elegir automáticamente el programa de menor coste, ya que podría aportar resultados de supervivencia menores. Para la comparación de alternativas tendremos que cuantificar la supervivencia (por ejemplo, en años de vida ganados) en ambas alternativas y comparar los costes por unidad de efecto (coste por año de vida ganado en cada alternativa). Este tipo de evaluación económica se denomina ***análisis coste-efectividad***.

Un análisis más sofisticado permitiría calcular la calidad de vida asociada a la supervivencia en los tratamientos del ejemplo anterior. Ello se obtiene mediante el ***análisis coste-utilidad***, que pretende medir la relación entre los costes de las intervenciones y los efectos de las mismas, a través de una medida que integra supervivencia y calidad de vida. Esta medida se denomina *utilidad*. Para su utilización como medida de resultado, la utilidad suele medirse en una escala entre 0 (muerte o peor estado de salud posible) y 1 (estado de perfecta salud), y se ajusta por la probabilidad de alcanzar esos resultados y por el periodo de tiempo esperado que estará afectado por un determinado valor de utilidad. Cuando se utilizan los años como medida de tiempo, estamos frente a un tipo de unidad denominada genéricamente *Años de Vida Ajustados por Calidad (AVACs)*.

Finalmente, en el ***análisis coste beneficio***, tanto los costes como los efectos de las intervenciones se miden en unidades monetarias. Este tipo de análisis tiene las ventajas de ofrecer la información sobre costes y resultados en las mismas unidades, y de permitir analizar el beneficio neto de cada alternativa, posibilitando el rechazo de aquellas cuyos costes sean mayores que los beneficios que aportan.

La siguiente tabla muestra la relación entre la valoración de costes y resultados y las alternativas.

Tabla. A6.3. Relación entre la valoración de costes y resultados. Alternativas

		¿Se examinan tanto los costes como los resultados?	
		NO	SÍ
¿Hay comparación entre dos o más alternativas?	NO	Se examinan sólo los resultados	Se examinan sólo los costes
		EVALUACIÓN PARCIAL	
		Descripción del resultado	Descripción del coste
		EVALUACIÓN PARCIAL	
	SÍ	Evaluación de la eficacia o de la efectividad	Análisis del coste
		EVALUACIÓN ECONÓMICA COMPLETA	
		1- Análisis de Minimización de Costes. 2- Análisis de Coste-Efectividad. 3- Análisis de Coste-Utilidad 4- Análisis de Coste-Beneficio.	

Fuente: Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes. Third edition*: Oxford: Oxford University Press; 2005.

4.2.3. Pruebas sobre la efectividad de las alternativas

Se trata de un punto central en la EE ya que, en general, existe poco interés en la provisión eficiente de servicios ineficaces o inefectivos. La EE es incapaz, por sí misma, de establecer la efectividad de las alternativas y, por ello, debe incluir los datos que avalen la eficacia/ efectividad de las alternativas consideradas. La evidencia sobre la eficacia de cada alternativa será más sólida si se basa en resultados de estudios experimentales. No obstante, en este supuesto conviene valorar si esta eficacia se mantendrá en las condiciones reales de aplicación.

4.2.4. Identificación y valoración de los costes relevantes

En principio, una EE debería incluir todos los costes de cada una de las alternativas, que pueden subdividirse en tres grupos básicos: 1) costes directos, ligados directamente al programa, sean de los servicios sanitarios o del paciente; 2) costes indirectos, ligados indirectamente al programa, que pueden provenir del paciente y sus familiares; y 3) costes externos al sistema sanitario (modificación de los sistemas de producción, distribución, etc.). La valoración de todos ellos debe hacerse, siempre que existan precios de mercado, en precios de mercado.

4.2.5. Consideración temporal de los costes y los beneficios

En general, los costes de los programas sanitarios suelen ser inmediatos, mientras que los efectos a veces se producen inmediatamente (intervención quirúrgica) o en un futuro más o menos lejano (detección precoz de la hipertensión arterial). Los individuos y los grupos muestran como norma una tasa de **preferencia temporal positiva** (prefieren disponer de

los recursos ahora en vez de en el futuro) y ello obliga a que la comparación entre costes y beneficios deba realizarse en un tiempo definido (habitualmente el presente). El ajuste temporal se realiza mediante la actualización o descuento del flujo futuro de costes y efectos-beneficios al momento actual.

4.2.6. Presentación de resultados

La presentación de resultados es un aspecto fundamental de los estudios de EE, ya que una presentación desagregada, además de la presentación de los índices sintéticos, facilita el examen crítico de los supuestos, así como la reproducibilidad del análisis y su aplicación a supuestos distintos. El principal índice para la presentación de resultados es el denominado **Ratio Coste Efectividad-Utilidad Incremental** $(C_1 - C_2) / (E_1 - E_2)$: cuando una alternativa es más costosa, pero también más efectiva que otra, es conveniente utilizar este índice para comparar los extras de costes y beneficios de pasar de una alternativa a otra.

4.2.7. Consideración del riesgo y la incertidumbre. Análisis de sensibilidad

Toda evaluación tiene incertidumbres, imprecisiones y problemas metodológicos. Un análisis depurado debería incluir lo siguiente: 1) identificar suposiciones metodológicas críticas; 2) identificar áreas de incertidumbre; y 3) emplear otras suposiciones o estimaciones para valorar la sensibilidad del análisis a las variaciones de las mismas. Esto se realiza mediante el análisis de sensibilidad, introduciendo cambios en las estimaciones menos fundamentadas y valorando la robustez que muestran los resultados del análisis a estos cambios. En general, se otorgará más confianza a los resultados originales cuando el análisis de sensibilidad no altere substancialmente las conclusiones del mismo.

4.2.8. Consideraciones distributivas y equidad

La EE está basada sobre criterios utilitaristas, valorando los costes y beneficios para el conjunto de la sociedad con independencia de sobre quién recaigan. Esta aproximación no siempre es adecuada en la atención sanitaria, que cumple una función redistributiva importante y puede conducir a asignaciones de recursos inequitativas. La EE es un instrumento importante de ayuda en la toma de decisiones, en la medida que explicita los costes y beneficios esperados de cada alternativa; pero no puede sustituir el propio proceso de decisión, ni la responsabilidad y los riesgos inherentes a la toma de decisiones, máxime cuando sus limitaciones pueden restringir de manera importante su utilización en una situación concreta y cuando la técnica no garantiza, en términos generales, que la alternativa más eficiente sea la más equitativa.

Literatura recomendada para conocer más

Técnicas de consenso:

Murphy MK, Black NA, Lamping DL, McNee CM, Sanderson CFB, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess.* 1998;2(3):1-100.

Campbell SM, Cantrill JA. Consensus methods in prescribing research. *J Clin Pharm Ther.* 2001;26(1):5-14.

Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ.* 1995;311(7001):376-80.

Olmos ME, Sánchez R, Venegas MA. Los consensos de expertos: una metodología útil en la toma de decisiones en salud. *Rev Colomb Cancerol.* 2006;10(1):50-60.

Black N. Consensus development methods. En: Pope C, Mays N. *Qualitative methods in health care.* Oxford: Blackwell Publishing; 2007. p 132-41.

Técnicas de investigación cualitativa:

Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess.* 1998;2(16):iii-ix,1-274.

Flick U. *An Introduction to qualitative research.* London: SAGE Publications; 1998.

García Ferrando M, Ibáñez J, Alvira F. *El análisis de la realidad social: métodos y técnicas de investigación.* Madrid: Alianza Editorial; 2007.

Calderón C, Fernández MJ. *Investigación cualitativa en atención primaria.* En: Martín A, Cano JF. *Atención primaria: conceptos, organización y práctica clínica.* 6ª ed. Madrid: Elsevier; 2008.

Pope C, Mays N. *Qualitative methods in health care.* Oxford: Blackwell Publishing; 2007.

Técnicas de investigación cuantitativa:

Eldridge S, Ashby D, Bennet C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ.* 2008;336(7649):876-80.

Browner WS, Newman TB, Cumming SR, Hulley SB. *Estimating simple size and power: the nitty-gritty en designing clinical research.* 2ª ed. Philadelphia: Lippincott Williams & Wilkins; 2001.

Motulsky H. *Intuitive statistics.* Oxford: Oxford University Press; 1995.

Técnicas de evaluación económica:

Guidelines for the economic evaluation of health technologies. 3ª ed. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.

Chiou C, Hay JW, Wallace JF, Bloom BS, Neuman PJ, Sullivan SD, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care.* 2003;41(1):32-44.

Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Métodos para la evaluación económica de los programas de asistencia sanitaria. Madrid: Díaz de Santos; 2001.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3ª ed. Oxford: Oxford University Press; 2005.

Ortún V. La economía en sanidad y medicina: instrumentos y limitaciones. Barcelona: Escola Universitària de Treball Social y la Llar del Llibre; 1990.

Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, Davies LM, Eastwood A. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess*. 2004;8(49):iii-iv,1-192.