

Métodos y herramientas para la **realización** de **ciclos** de **mejora** de la **calidad** en **servicios** de **salud**

Pedro J. Saturno Hernández



Instituto Nacional
de Salud Pública



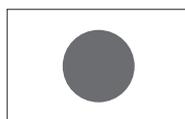
Métodos y herramientas
para la realización de ciclos de mejora
de la calidad en servicios de salud

Métodos y herramientas para la **realización** de **ciclos** de **mejora** de la **calidad** en **servicios** de **salud**

Pedro J. Saturno Hernández



Instituto Nacional
de Salud Pública



Métodos y herramientas para la realización de ciclos de mejora de la calidad en servicios de salud

Primera edición, 2015

D.R. © Instituto Nacional de Salud Pública
Av. Universidad 655, Col. Santa María Ahuacatlán
62100 Cuernavaca, Morelos, México

ISBN: 978-607-511-152-0

Impreso y hecho en México
Printed and made in Mexico

Citación sugerida: Saturno-Hernández PJ. Métodos y herramientas para la realización de ciclos de mejora de la calidad en servicios de salud. Cuernavaca, México: Instituto Nacional de Salud Pública, 2015.

Acerca del autor

Pedro J. Saturno Hernández es médico, doctor en Salud Pública por la Universidad de Harvard, profesor honorífico de Salud Pública de la Facultad de Medicina de la Universidad de Murcia (España) y científico visitante y miembro del Grupo de Salud Global de la Escuela de Salud Pública de la Universidad de Harvard. Dirige una Maestría en línea sobre Gestión de la Calidad en Servicios de Salud, que actualmente se imparte en su 15° edición en la Universidad de Murcia, en su segunda en el Instituto Nacional de Salud Pública (INSP), en México y en su primera en la Universidad de Rio Grande del Norte, en Brasil.

Previamente se ha desempeñado como director general adjunto de Planificación Sanitaria y Educación en el Ministerio de Salud de España. Tiene una extensa experiencia como consultor sobre gestión de la calidad en servicios de salud para agencias internacionales como la Organización Mundial de la Salud (OMS) y la Agencia de Estados Unidos para el Desarrollo Internacional (USAID), entre otras. Ha participado en varios grupos de trabajo y comités nacionales e internacionales en relación con la gestión de la calidad, prevención en salud y educación a lo largo de los últimos 30 años, ha impartido cursos especializados y conferencias en numerosos ámbitos y ha ayudado a la implantación de programas de mejora de la calidad en varios países e instituciones.

Actualmente es AXA Professor on Healthcare Quality y director adjunto del Centro de Investigación en Calidad y Encuestas del INSP.

Contenido

Presentación	11
Capítulo 1.	
Marco conceptual para la gestión de la calidad	13
Capítulo 2.	
Identificación y priorización de oportunidades de mejora	39
Capítulo 3.	
Métodos y análisis de los problemas de calidad	61
Capítulo 4.	
La construcción de criterios para evaluar la calidad	81
Capítulo 5.	
Diseño de estudios de nivel de calidad y la investigación de causas.	
Componentes de una evaluación	101
Capítulo 6.	
Diseño de estudios para la investigación de causas	123
Capítulo 7.	
Análisis y presentación de los datos de una evaluación	141
Capítulo 8.	
Diseño de intervenciones para mejorar	167
Capítulo 9.	
Implementación de intervenciones para la mejora de la calidad.	
Métodos de seguimiento y control	179
Capítulo 10.	
Cierre del ciclo de mejora. Reevaluación, análisis y presentación de resultados comparativos de dos evaluaciones	189

Presentación

La mejora de la calidad de los servicios de salud precisa una atención específica, continua y practicada con una metodología rigurosa en la que tienen cabida la innovación y la experimentación en formas de hacer las cosas, pero que siempre ha de venir acompañada de medición. Lo que no se mide no se conoce bien, lo que no se conoce no se puede mejorar, y la mejora conseguida sólo puede tenerse como cierta si se documenta adecuadamente.

Sin apartarse de este principio general, se puede entrar a la dinámica de la mejora continua de la calidad con tres enfoques o grupos de actividades, diferenciables metodológicamente pero en definitiva complementarios: (i) planificación o diseño de la calidad, anticipándose a los problemas y diseñando los procesos para conseguir –cuando se implementen– los resultados deseados; (ii) monitorización de indicadores válidos, confiables y útiles para identificar –cuando se midan– si estamos en los niveles de calidad deseados, y (iii) ciclos de mejora sobre problemas identificados en la implementación de los procesos, resultados obtenidos o resultados de la monitorización. Los tres enfoques son relevantes; los tres son una buena puerta de entrada para la mejora de la calidad; los tres se interrelacionan, pero el central es el de ciclos de mejora. Sin el complemento de este enfoque, la monitorización y planificación pueden ser inútiles o, como mínimo, insuficientes para gestionar y mejorar la calidad. En los tres, sin embargo, es imprescindible un bagaje metodológico adecuado para la implementación rigurosa y eficiente.

El presente manual aporta los conocimientos teóricos y, sobre todo, prácticos para la realización paulatina de un ciclo de mejora sobre los problemas de calidad comenzando por su identificación y definición y pasando por todas las etapas analíticas, de intervención y de medición que culminan con la documentación de la mejora conseguida. Complementa otros dos volúmenes: uno, ya publicado, se centra en la metodología para la monitorización;¹ el segundo, en preparación, aborda los métodos y herramientas para el diseño o planificación de la calidad en servicios de salud. En conjunto, estos manuales constituyen un material de apoyo y referencia práctica para hacer realidad el compromiso con la mejora de la calidad, que recae en todos los elementos, personas y niveles del sistema de salud.

Hace ya años que se reconoce la responsabilidad que, además de los dirigentes del sistema, ejerce el propio personal de salud en la evaluación y mejora de la calidad, aunque sólo sea desde el punto de vista de la ética profesional. Se advierte como absolutamente necesario que la capaci-

1. Saturno-Hernández PJ. Métodos y herramientas para la monitorización de la calidad en servicios de salud. Cuernavaca, México: Instituto Nacional de Salud Pública, 2015.

tación en la metodología adecuada para la mejora de la calidad forme parte de los conocimientos y habilidades de todo el personal de salud.²⁻⁵

Existe un gran vacío, no sólo en los programas de formación de pregrado y posgrado, sino también en el sector salud. No hay manuales que, como el que aquí se presenta, contribuyan con la consecución de este objetivo ético y profesional para poder avanzar en la necesaria capacidad organizacional del sistema de salud con el fin de ofrecer servicios de la mayor calidad posible.^{6,7}

La forma secuencial en que están organizados los contenidos permite utilizar este manual como guía metodológica para cada uno de los pasos de un ciclo de mejora, de forma que se provee al lector de las herramientas esenciales para aplicarlas de forma específica en cada uno de los pasos o etapas.

Sin dejar de lado elementos conceptuales o teóricos, el enfoque es eminentemente práctico. El presente manual, además, está avalado por más de quince años de utilización en la realidad de los servicios de salud, durante los cuales ha formado parte del material de apoyo de la Maestría en Gestión de Calidad en los Servicios de Salud, cuya metodología está orientada a la solución de problemas reales y que actualmente se ofrece en España, México y Brasil.

Por otra parte, la estructura del manual sigue y pone énfasis en los aspectos requeridos a nivel internacional para poder reportar un ciclo de mejora en términos publicables en revistas científicas.⁸ De esta forma, no sólo es una guía para realizar las actividades del ciclo de mejora con el rigor necesario, sino también para concretarlo sin perder de vista el tipo de información que ha de ser útil para comunicar la experiencia y que otros profesionales y unidades puedan aprender de ella. Con esto, se pretende incidir en la extensión y conocimiento de las herramientas para la mejora continua mostrando su utilidad y beneficios, tanto para los profesionales, como para el sistema en su conjunto y, principalmente, para la población a la que se atiende.

-
2. Berwick DM, Enthoven A, Bunker JP. Quality management in the NHS: the doctor's role. *BMJ* 1992;304:235-239.
 3. Brennan TA. Physicians' professional responsibility to improve the quality of care. *Academic Medicine* 2002;77(10):973-980.
 4. Peterson LE, Jaen CR, Phillips RL. Family physician participation in quality improvement. *J Am Board Fam Med* 2013;26:626-627.
 5. Jones AC, Shipman SA, Ogrinc G. Key characteristics of successful quality improvement curricula in physician education: a realist review. *BMJ Qual Saf* Published Online First: 30 Sep. 2014. doi:10.1136/bmjqs-2014-002846.
 6. Jones AC, Shipman SA, Ogrinc G. Key characteristics of successful quality improvement curricula in physician education: a realist review. *BMJ Qual Saf* Published Online First: 30 Sep. 2014. doi:10.1136/bmjqs-2014-002846.
 7. Therman S, Massoud R, Saturno P. Quality of care. A process for making strategic choices in health systems. Ginebra: WHO, 2006.
 8. Ogrinc G, Davies L, Goodman D, et al. SQUIRE 2.0 (Standards for quality improvement reporting excellence): revised publication guidelines from a detailed consensus process *BMJ Qual Saf* Published Online First: 14 Sep.2015 doi:10.1136/bmjqs-2015-004411.

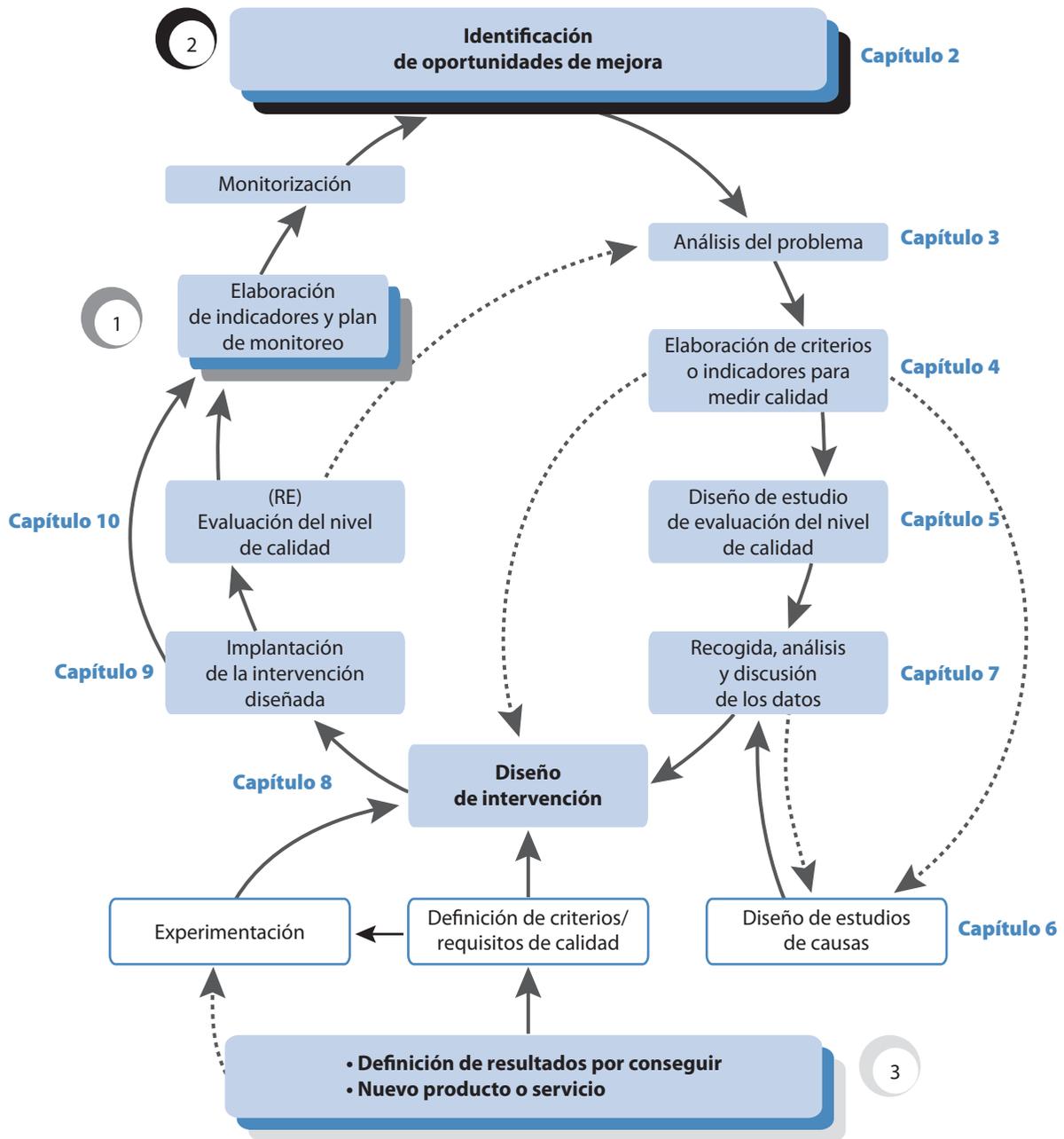
Capítulo 1

Marco conceptual para la gestión de la calidad

Contenido general y objetivos específicos

1. El concepto general y operativo de la calidad
2. Satisfacción del profesional: concepto de cliente interno
3. Concepto de calidad total
4. Gestión de la calidad. Componentes esenciales de un sistema de gestión de la calidad
5. Los bloques estratégicos para la mejora de la calidad
6. Las actividades para la mejora de la calidad
7. Programas internos y externos en la estrategia para la mejora de la calidad
8. Responsabilidades en la mejora de la calidad
9. Estrategias de implantación de actividades. El ciclo de mejora

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Puntos de partida

Vía de acción normal

Vía contingente o alternativa

1 Monitorización

2 Ciclo de mejora

3 Diseño

Contenido general

En este capítulo se da un marco de referencia global para las actividades de gestión de la calidad. Se aborda la definición de calidad, necesaria para dar sentido (validez) a los aspectos que vamos a medir y mejorar, y se especifican las principales actividades y componentes que ha de tener un sistema de gestión de la calidad, así como responsabilidades y estrategias de implantación. Finalmente, se describen de forma introductoria los distintos pasos de un ciclo de mejora, la actividad central a cuya realización está dirigido el presente manual metodológico.

Objetivos específicos

1. Diferenciar definiciones genéricas de calidad, aplicables a cualquier tipo de producto o servicio, de las definiciones específicas de actividades y de productos en el ámbito de los servicios de salud.
2. Describir las dimensiones o aspectos generales que entraña el concepto de calidad de los servicios de salud.
3. Distinguir los conceptos de cliente interno y cliente externo, aplicados a los servicios de salud.
4. Identificar los componentes esenciales de un programa o sistema de gestión de la calidad.
5. Identificar los grandes bloques estratégicos para la mejora de la calidad.
6. Distinguir los tres grandes grupos de actividades que componen los programas de gestión de la calidad.
7. Describir los puntos de partida (inicio) y los objetivos directos de cada grupo de actividades.
8. Explicar las diferencias entre los enfoques interno y externo.
9. Identificar el tipo de enfoque que corresponde a cada uno de los protagonistas (administración, personal de atención directa, pacientes/población) y las estrategias para la mejora de la calidad en los servicios de salud.
10. Justificar la responsabilidad y el protagonismo que corresponde al personal que presta directamente los servicios, como proveedor y como cliente interno del sistema.
11. Razonar las ventajas e inconvenientes de las tres principales formas de comenzar las actividades para la mejora de la calidad.
12. Describir los principales pasos o fases de un ciclo de mejora.

“Cuando la mente se ensancha para dar cabida a una idea nueva nunca recupera su dimensión original”.

1. El concepto general y operativo de la calidad

Definir lo que se entiende por calidad de la atención no es una disquisición teórica irrelevante para la práctica, si lo que pretendemos es explicitar el marco de referencia para su evaluación y mejora. Las más elementales reglas de validez exigen tener claro qué queremos decir con *calidad de la atención* cuando va a ser ese concepto el que vamos a cuanti-

ficar. Vamos a necesitar definir para medir y medir para saber cómo podemos intervenir para mejorar aquello que previamente hemos definido.

Para orientarnos en la concreción de lo que vamos a entender por calidad, es conveniente distinguir tres niveles, de más general a más específico, tal como está representado en la figura 1.1. Los tres niveles de conceptualización han de ser consistentes entre sí. Vamos a ver a qué se refiere cada uno de ellos.

Necesitamos definir qué es calidad para, entre otras cosas, dar validez a su medición y mejora.

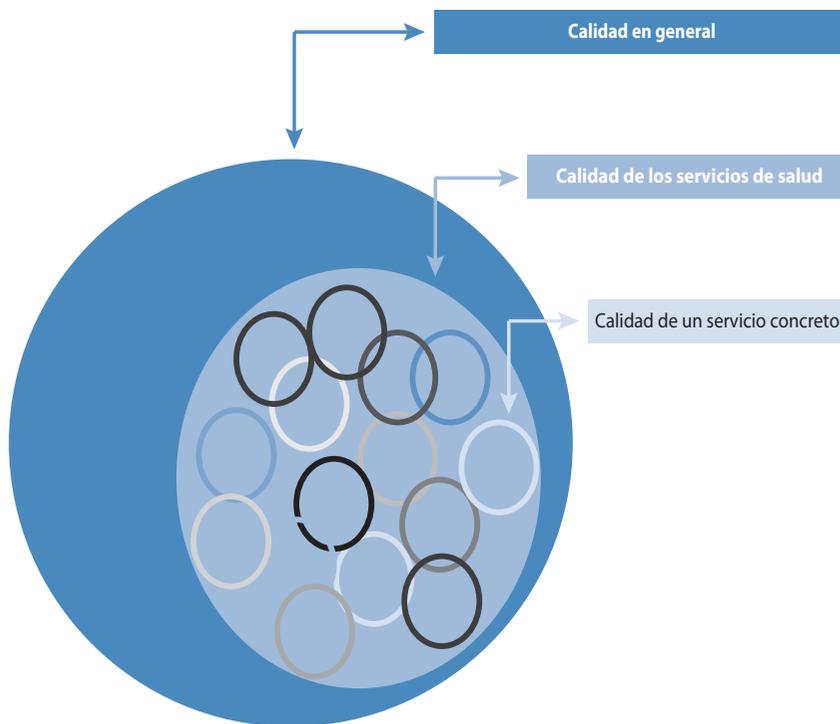


Figura 1.1. Los niveles del concepto de calidad

En general, la definición de calidad de un producto o servicio ha de hacerse en función de las necesidades y expectativas de quien lo recibe.

1.1. El concepto general de calidad

De una manera amplia, el concepto de calidad puede aplicarse a muy diversos aspectos de una determinada actividad industrial o de servicios: desde niveles filosóficos globales, hasta la calidad de los procesos, la organización del personal, los objetivos de la empresa, etc. Aquí nos vamos a referir de momento a calidad del producto o servicio que se ofrece. La pregunta por responder sería: ¿Cómo sabremos si un producto o servicio es de calidad? Como ejemplo, veamos las respuestas de algunos de los autores más influyentes sobre la calidad de la industria: Juran, Deming e Ishikawa. Juran define la calidad como “idoneidad o aptitud para el uso”. Un producto o servicio será de calidad si sirve para lo que está previsto que sirva. La aplicación de este concepto a un servicio o producto concreto exige una definición previa de qué necesidades y expectativas está previsto que sean satisfechas con el producto o servicio que se va a evaluar, las cuales serían diferentes según el producto o servicio y determinarán las características que éste ha de tener para que sea de calidad.

Deming propone definir la calidad “en función del sujeto que ha de juzgar el producto o servicio”. La idea básica es la misma: quién compra o demanda el servicio lo hace en función de unas necesidades y expectativas que son las que van a determinar si lo que recibe es o no de calidad.

Ishikawa utiliza la misma idea: la calidad del producto o servicio es la “satisfacción de los requisitos de los consumidores de ese producto o servicio”. Hay que determinar, pues, cuáles son esos requisitos para poder saber si el producto o servicio es de calidad. Como puede apreciarse, hay un esquema básico subyacente a estas definiciones, que nos puede servir para describir y analizar la calidad de cualquier producto o servicio. Este esquema básico parte de tres componentes que hay que identificar y caracterizar: proveedor, servicio o producto que se ofrece y receptor del servicio o producto. Para que haya calidad, el que ofrece el servicio ha de tener en cuenta las necesidades y expectativas del receptor a quién va dirigido, y modelar el servicio o producto de ma-

nera que sintonicen perfectamente con éstas. La falta de sintonización, por desconocimiento de las necesidades y expectativas de los receptores o por no ser capaces de diseñar los servicios de forma que satisfagan estas necesidades y expectativas, serán defectos de calidad. Este esquema básico está representado en la figura 1.2. En consecuencia, la calidad va a depender de lo correcta que sea la interpretación que hagamos de las necesidades y expectativas de los usuarios y de cómo las incorporemos al diseño de los productos o servicios.

1.2. El concepto de calidad en servicios de salud

Hay múltiples definiciones de lo que se entiende por calidad de los servicios de salud. Todas responden a diversas interpretaciones sobre qué significa atender las necesidades sanitarias de la población a la que se ofrecen los servicios cuya calidad queremos definir.

En la definición adoptada en el Programa Ibérico (Saturno PJ, Imperatori E, Corbella A, 1990), calidad de la atención es “la provisión de servicios accesibles y equitativos, con un

El concepto de calidad y el contexto de los centros sanitarios van a ser la base para las actividades de gestión de la calidad

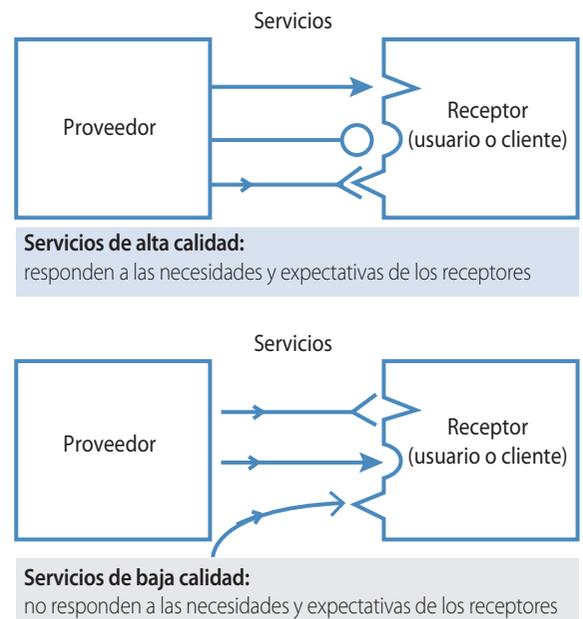


Figura 1.2. Esquema para la definición de calidad

nivel profesional óptimo, que tiene en cuenta los recursos disponibles y logra la adhesión y satisfacción del usuario". Esta definición, como cualquier otra que decidamos adoptar, explicita una serie de dimensiones que nos dan pistas de hacia dónde mirar para valorar si hay o no calidad en los servicios que se van a valorar. Al existir una gran variedad de definiciones, existe también una apreciable variedad de esquemas de dimensiones, factores, componentes o atributos (de todas estas formas se les llama). El cuadro 1.1 contiene algunos de estos esquemas. El lector interesado puede consultar la bibliografía que se cita en este capítulo. A continuación vamos a revisar el significado de las dimensiones más habituales (calidad científico técnica, efectividad, eficiencia, accesibilidad, satisfacción, adecuación, continuidad y seguridad) para terminar

proponiendo un esquema de sólo cuatro, normalmente suficiente como marco de referencia, y que resulta acorde con la definición que hemos propuesto.

La competencia profesional o *calidad científico-técnica* de la atención es la dimensión central y menos discutible de la calidad. Hace referencia a la capacidad de los proveedores de utilizar el más avanzado nivel de conocimientos existente para abordar los problemas de salud, es decir, para producir salud y satisfacción en los usuarios. En la mayoría de las definiciones, incluye tanto los aspectos científico-técnicos, en sentido estricto, como los de trato interpersonal, tal como es entendida desde la conceptualización de los componentes de la calidad que hizo Donabedian, quien también la situó en el centro de lo que se podría entender por calidad en servicios de salud.

Cuadro 1.1
Dimensiones de la calidad de la atención según diversos esquemas

Vuori (componentes) (1982)	Palmer (1983)	Nutting et al. (atributos) (1987)	Joint Comm. (factores) (1989)	Joint Comm. (componentes) (1990)	Programa Ibérico (1990)	IOM (2001)	WHO (2006)
Calidad científico-técnica	Competencia profesional*	Competencia profesional	—	—	Calidad científico-técnica	—	—
—	Accesibilidad/* Equidad	Distribución de la asistencia/ Integralidad	Accesibilidad	Accesibilidad/ Disponibilidad	Accesibilidad/ Equidad	Equidad	Accesibilidad/ Equidad
—	Aceptabilidad/* Satisfacción	—	Participación	Aceptabilidad	Satisfacción	Servicio centrado en el paciente	Aceptabilidad centrada en el paciente
Efectividad	Efectividad	Efectividad	Efectividad	Efectividad	—	Efectividad	Efectividad
Eficiencia	Eficiencia	Eficiencia	Eficiencia	Eficiencia	—	Eficiencia	Eficiencia
Adecuación	—	—	Adecuación	Adecuación	—	—	—
—	—	Continuidad/ Coordinación	Continuidad	Continuidad	—	—	—
—	—	—	Eficacia	—	—	—	—
—	—	—	Oportunidad temporal	—	—	Oportunidad	Accesibilidad
—	—	—	Privacidad	—	—	—	—
—	—	—	Confidencialidad	—	—	—	—
—	—	—	Seguridad	—	—	Seguridad	Seguridad
—	—	—	Apoyo estructural	—	—	—	—

*: Dimensiones mantenidas en publicaciones posteriores

Sin lugar a dudas, es la dimensión que mejor se entiende y más frecuentemente medida como representante de la calidad de los servicios de salud; significa atender de forma científica y humana las necesidades de salud.

Efectividad es la medida en que una determinada práctica o atención sanitaria mejora el estado de salud de la población concreta que la recibe. Durante algún tiempo ha venido utilizándose de forma confusa o intercambiable con el término eficacia. Sin embargo, en la actualidad, de acuerdo con las propuestas de K. L. White refrendadas por la OTA (Office of Technology Assessment) estadounidense, por eficacia se entiende “la probabilidad de beneficio de una determinada tecnología en condiciones ideales de uso”, mientras que efectividad sería “la probabilidad de beneficio de una determinada tecnología en condiciones normales de uso”. Lo más frecuente al definir calidad es la referencia a la efectividad de la atención.

Eficiencia se define en términos de la relación entre coste y producto es decir, un máximo de efectividad o unidades de producto, dado un determinado coste, o un mínimo coste dadas unas determinadas exigencias de efectividad o unidades de producto. Por supuesto, se asume que una mayor eficiencia se asocia a una mayor calidad.

La *accesibilidad* puede definirse como la facilidad con que la atención sanitaria puede obtenerse en relación con los aspectos (barreras) organizacionales, económicos, culturales y emocionales. Esta dimensión no se incluye en todas las definiciones de calidad. También es cierto que aquellas que la incluyen a veces introducen matices muy diferentes en su conceptualización operativa, medible. Palmer incluye en ella el concepto de equidad: accesibilidad es “la medida en que los cuidados apropiados son recibidos por la población de forma equitativa” o “la distribución equitativa y oportuna, a tiempo, de la atención sanitaria apropiada, para aquéllos con necesidades equivalentes”. En todo caso, lo que subyace es la necesidad de cuantificar si la atención sanitaria llega o no a quien la necesita y cuando la necesita.

Un matiz importante que podemos introducir es aplicar el concepto de accesibilidad no sólo a la obtención de servicios iniciales o contacto principal con el sistema de salud, sino también a la existencia o inexistencia de barreras organizativas, culturales o de otro tipo una vez que se ha accedido al sistema de salud. Podríamos llamar este segundo aspecto accesibilidad interna y englobaría toda la problemática de los tiempos de espera, descoordinaciones entre profesionales y niveles asistenciales, etcétera, que impiden la fluidez normal que debería tener un proceso asistencial visto en su totalidad. A este tipo de accesibilidad se refiere la dimensión oportunidad del más reciente esquema del IOM.

La *satisfacción* puede definirse como la medida en que la atención sanitaria y el estado de salud resultante cumplen con las expectativas del usuario. Según las distintas definiciones de calidad, esta dimensión se incluye o se confunde con el concepto más amplio de aceptabilidad o el concepto más reciente de atención centrada en el paciente, ampliado también a familia y población, del cual la satisfacción sería uno de los resultados. En todo caso, así se llame satisfacción o aceptabilidad, de lo que se trata es de tener en cuenta la manera en que la atención es percibida y protagonizada por el usuario y sus reacciones como consecuencia de ello.

La *adecuación* (*appropriateness*) es otra de las dimensiones mencionadas con frecuencia, pero de la que no existe una definición operativa uniforme. En la lista de la JCAHO (Joint Commission on Accreditation of Healthcare Organizations) estadounidense, la adecuación es la medida en que la atención se corresponde con las necesidades del paciente; es decir, adecuado es entendido como sinónimo de correcto, conveniente o necesario para el problema concreto que es atendido. Éste es un concepto que parece muy difícil de separar, a la hora de ser medido, del concepto de calidad científico-técnica o competencia profesional. Un matiz diferente se introduce en la definición de dimensiones de OMS (2006), en la que la adecuación de competencia técnica y recursos para las necesidades de salud, queda incluido en la dimensión de accesibilidad.

La *continuidad* es, con toda probabilidad, la dimensión de la calidad de la atención más compleja de conceptualizar y medir. De entrada, ya resulta inquietante que sea la dimensión sobre la que existen más versiones diferentes. La conceptualización más prevalente y atractiva, parece ser que es la que propuso Shortell: “la medida en que la atención necesaria se provee de una forma ininterrumpida y coordinada”, en cierto modo equivalente a lo que hemos llamado *accesibilidad interna*.

La *seguridad* es la última de las dimensiones que se ha hecho explícita en las definiciones de la calidad en los servicios de salud. Su prominencia actual se basa en los datos que se hicieron públicos primero en Estados Unidos y después en muchos otros países, los cuales demostraron en qué medida los servicios de salud pueden ser la causa directa de daño evitable en el paciente. De ahí la definición inicial de la seguridad como la “ausencia de lesiones a causa de la atención sanitaria que se supone puede ser beneficiosa” (IOM, 2001). Obviamente es una dimensión íntimamente ligada a la calidad científico-técnica, y más aún en la definición más aceptada en la actualidad, según la cual la seguridad no es sólo la ausencia de lesiones sino la “ausencia de daño innecesario real o potencial asociado a la atención sanitaria” (OMS, 2008). Es decir: incluye las situaciones de riesgo innecesariamente aumentado

de que se produzca daño, aunque éste no llegue a producirse. Por tanto, el énfasis no está en que la atención sanitaria sea más o menos efectiva o más o menos beneficiosa, sino en su potencialidad para producir daño innecesario.

En el cuadro 1.1 se incluyen esquemas que contienen dimensiones, atributos o factores de la calidad diferentes a los comentados hasta aquí. Revisar cada uno de ellos sería muy prolijo, máxime cuando pensamos que lo más lógico es intentar simplificar. Destacamos, sin embargo, el del IOM, cuyas definiciones pueden verse en el cuadro 1.2, por la repercusión que tuvo y sigue teniendo como esquema de base para la búsqueda y clasificación de indicadores para monitorizar la calidad de los sistemas de salud. Un esquema muy parecido es el que contiene la publicación de la OMS (2006) citada en la bibliografía, sólo que con referencia explícita al nivel comunitario e incluyendo la dimensión *accesibilidad*.

¿Se puede simplificar? Las cuatro dimensiones clave y una agrupación con interés práctico.

Un esquema reducido de dimensiones que podemos adoptar se reduce a cuatro: calidad científico-técnica o competencia profesional, atención centrada en el paciente/familia/población, accesibilidad/equidad y seguridad. ¿Por qué puede prescindirse de explicitar

Proponemos un esquema simplificado de cuatro dimensiones: calidad científico-técnica, accesibilidad, atención centrada en el paciente, y seguridad.

Cuadro 1.2

Las seis dimensiones clave de la calidad* (IOM, 2001)

● Seguridad	Evitar lesiones a causa de la atención sanitaria, que se supone debe ser beneficiosa.
● Efectividad	Proveer servicios basados en el conocimiento científico a todos los que pueden beneficiarse de ellos, y no proveerlos a aquellos que probablemente no se beneficien de ellos (evitar sub y sobre utilización respectivamente).
● Servicio centrado en el paciente	Proveer atención sanitaria que es respetuosa con el paciente y que responde a sus preferencias individuales, necesidades y valores, asegurando que éstos guíen todas las decisiones clínicas.
● Oportunidad	Reducir las esperas y, a veces perjudiciales, retrasos tanto para los que reciben como para los que proveen la atención sanitaria.
● Eficiencia	Evitar desperdicio de equipamiento, insumos, ideas y energía.
● Equidad	Proveer servicios que no varían en calidad según características personales tales como género, etnia, localización geográfica y estatus socioeconómico.

Adaptado de: Institute of Medicine: Crossing the Quality Chasm. A New Health System for the 21st Century. National Academy Press, Washington 2001

* Referidas fundamentalmente a la atención sanitaria individual, no a las poblaciones o comunidades

efectividad, eficiencia, adecuación y continuidad, y de otras que figuran en el cuadro 1.1 como dimensiones de medición directa de la calidad? No es porque se considere que estas dimensiones no puedan estar directamente relacionadas con la calidad de la atención. Las razones fundamentales tienen que ver con la práctica de las evaluaciones sobre calidad de la atención, si se trata de medirla para determinar qué hacer para mejorarla, y evitar en la medida de lo posible los solapamientos que existen entre ellas.

Tomemos, por ejemplo, la *efectividad* y la *eficiencia*. Desde luego que más efectividad y más eficiencia son más calidad; pero, efectividad y eficiencia ¿de qué?, provocadas o decididas ¿por quién? La efectividad y la eficiencia son atributos de las acciones diagnósticas y terapéuticas, en sentido amplio, que los proveedores de los servicios han decidido utilizar o poner en marcha para la resolución de un determinado problema de salud. Depende de la competencia profesional del proveedor haber seleccionado y ejecutado las acciones más efectivas y eficientes que convenían al problema afrontado, con las limitaciones de los recursos disponibles. Es decir, al medir efectividad y eficiencia medimos la calidad científico-técnica o la competencia profesional para actuar de forma adecuada. Serían así indicadores de resultado de la calidad científico-técnica.

En relación con otras dimensiones finalmente no incluidas en nuestra propuesta ya hemos argumentado antes lo difícil que resulta considerar la adecuación de forma separada de la calidad científico-técnica. Si consideramos la continuidad, de indiscutible relevancia a la hora de medirla para mejorar la calidad, sus diversos componentes o (sub)dimensiones son identificables en su mayoría como cuestiones que miden, en realidad, accesibilidad, atención centrada en el paciente o competencia profesional, con mayor o menor importancia relativa, según el problema de salud que estemos evaluando y el aspecto de la continuidad que valoremos.

La *seguridad* merece mención aparte. Aunque los argumentos al respecto son similares a los expuestos para la adecuación y continuidad, en el sentido de que la seguridad depende de una mezcla de factores relacionados con la

calidad científico-técnica (de los diversos profesionales y niveles del sistema de salud) y la accesibilidad, el auge que está teniendo en los últimos años la atención específica a este aspecto de la calidad asistencial con base en la magnitud de sus deficiencias en el sistema de salud, podría justificar su consideración particular a efectos de medición y mejora.

Finalmente, conviene señalar que, del conjunto de dimensiones de la calidad, hay algunas que se miden y aplican sobre todo a la atención clínica individual de aquéllos que acceden a la atención (calidad científico-técnica, seguridad, atención centrada en el paciente); mientras que otras implican un ámbito comunitario o poblacional (accesibilidad/equidad, atención centrada en la población). En algunos ámbitos y momentos del desarrollo histórico de la calidad en los servicios de salud, ha primado exclusivamente la consideración de las dimensiones de atención clínica individual, pero conviene no olvidar la aplicación del concepto en sus dimensiones poblacionales y de sistema.

1.3. Calidad de un determinado proceso o servicio: el tercer nivel del concepto de calidad

A pesar de la claridad con que hayamos entendido y asumido lo expuesto hasta aquí, hay que dar un paso más para hacer operativo el concepto de calidad. El entender y adoptar conscientemente una definición de calidad de los servicios de salud, de la que podemos deducir unas dimensiones que interpretan lo que pensamos que da satisfacción a las necesidades y expectativas de la población atendida, nos confiere un importante marco de referencia, pero es aún insuficiente como instrumento para la medición directa de la calidad. Son muchos los servicios diferentes que ofrece cualquier establecimiento sanitario y es necesario traducir estas dimensiones generales a requisitos concretos para cada servicio, problema de salud o tipo de paciente, de forma que podamos valorar si es o no de calidad. En cada tipo de paciente y problema de salud la importancia relativa de cada dimensión puede ser diferente, así como los requisitos de calidad que las caracterizan. Piénsese, por ejemplo, lo muy diferente que han de ser los requisitos

La definición general de calidad de los servicios de salud es aún insuficiente para medir. Tenemos que especificar cómo se traducen las dimensiones generales en requisitos de calidad de cada producto o servicio que se ofrece.

de calidad científico-técnica para las distintas enfermedades crónicas, en comparación con procedimientos agudos, autolimitados, o con problemas de tipo preventivo. Por eso, vamos a necesitar de la definición de los requisitos específicos de cada servicio o tipo de paciente para poder, finalmente, cuantificar la calidad. Este tercer nivel de especificación es el que resulta realmente operativo y, por lo tanto, el que vamos a manejar de manera preferente para su evaluación y mejora. La definición de calidad, de esta forma, es diferente y peculiar para cada servicio y se puede llegar a ella intentando responder las siguientes preguntas: para un servicio concreto (atención al hipertenso, a la diabetes, una determinada intervención quirúrgica, la resolución de una solicitud de cita para consulta, etc.) ¿qué quiere decir que sea accesible?, ¿cómo me aseguro de que no tiene problemas de accesibilidad?, ¿qué quiere decir que se ofrece con un nivel óptimo de calidad científico-técnica?, ¿qué características me indican este nivel?, ¿qué importancia tiene la aceptabilidad y satisfacción de los usuarios?, ¿cómo puedo saber si están presentes? Una respuesta exhaustiva a todas estas preguntas nos producirá un listado de características o requisitos que nos definirían una atención de calidad para el problema de salud que intentamos definir, y que nos permitirán medirla.

2. Satisfacción del profesional: concepto de cliente interno

La satisfacción del profesional no ha sido mencionada en las definiciones de calidad revisadas, ni es frecuente encontrarla de forma explícita en ninguna definición. ¿Por qué? La razón principal, si recordamos el esquema básico (proveedor → servicio → receptor) (figura 1.2), es que el producto o servicio que hemos intentado definir hasta ahora es aquél que recibe el usuario o cliente externo al proveedor, entendiendo por proveedor al individuo, la organización o el sistema que ofrece el servicio. Adicionalmente, hemos aclarado que íbamos a hablar de calidad del producto o servicio, y no de otras *calidades* como puede ser la de la organización, objetivos, personal, etcétera. La satisfacción del profesional ha de ser contemplada dentro de la *calidad* de la organización,

el componente proveedor, pero no es un elemento de la calidad del servicio en sí, aunque puede influir en él, al igual que otros muchos factores del sistema o de la organización. Vamos a detenernos un poco más en detallar estos dos aspectos: (i) cómo influye la satisfacción del profesional en la calidad del servicio y (ii) su relevancia como un elemento interno del sistema.

2.1. La satisfacción del profesional y la calidad del servicio prestado

La justificación fundamental implícita y subyacente para no incluir la satisfacción del profesional en la definición de la calidad del servicio que se ofrece al usuario externo es que se trata de un elemento del sistema y no del servicio ofrecido. La satisfacción del profesional puede considerarse como un criterio de la calidad de la organización o institución, un rasgo estructural que puede *favorecer* el buen funcionamiento y la calidad del servicio.

Adicionalmente, la evidencia empírica sostiene que, al igual que en la industria, la satisfacción del profesional puede influir en un mejor trabajo, así como el que una percepción del trabajo bien hecho produce satisfacción en los profesionales, pero no determina la calidad del servicio.

La relación satisfacción del profesional → calidad para el cliente externo no es siempre la misma, ni lineal, ni consistentemente importante. Lo que parece más claro es una relación de la satisfacción con la productividad y la estabilidad o rotación en los puestos de trabajo.

2.2. La satisfacción del profesional como cliente interno

Los individuos y los departamentos o cualesquiera unidades organizativas que integran un determinado centro asistencial son, a su vez, proveedores y clientes unos de otros, aunque sus actividades deban estar interconectadas con el fin exclusivo de servir a los clientes externos. Este concepto de *clientes internos* que reciben servicios para la provisión de un determinado producto final, lleva a incluir la satisfacción de los profesionales como parte de la definición de calidad de los servicios internos o, al menos,

La satisfacción del profesional forma parte de la calidad del sistema u organización, no del servicio que se ofrece, aunque pueda influir en él.

como uno de sus determinantes de importancia y, desde luego, como un factor relevante para el funcionamiento del centro. Un *cliente interno* sería aquél que recibe servicios de otros elementos del sistema (radiografías, análisis, servicios administrativos, etc.), y su satisfacción a la hora de definir la calidad de ese servicio interno es tan importante como para la definición de la calidad de los servicios externos.

Aparte de eso, aunque pueda argumentarse que la satisfacción de los profesionales influya en la calidad de su trabajo para el cliente externo y viceversa, es muy arriesgado decir que, al medir la satisfacción del profesional, estamos midiendo calidad de la atención al cliente externo. Sí estaremos midiendo, en cambio, un aspecto de la calidad de la organización y de sus servicios internos.

Calidad total implica una referencia a la calidad del producto o servicio y también a la calidad de la organización y sus relaciones con otras organizaciones.

3. Concepto de calidad total

El adjetivo *total* cualificando al concepto *calidad* es algo de uso relativamente prevalente hoy en día en los medios sanitarios. En la literatura anglosajona, en dónde se ha originado el término, el adjetivo *total* unido a *calidad* se empezó a utilizar en los años cincuenta y se sigue utilizando para adjetivar, no siempre a la palabra *calidad*, sino también a una tercera palabra, que es en realidad la que se conceptualiza de una forma particular el añadirle los dos términos *total* y *calidad*. Así, encontramos *total quality control*, *total quality system* y *total quality management*. ¿Qué se quiere significar al añadir la palabra *total*? Si vamos a las fuentes originales de los inventores del concepto, queda claro que lo que introduce es un importantísimo matiz de ámbito de actuación al control de calidad, al sistema de calidad y a la gestión de calidad: que el ámbito en el que actúen sea *total* quiere decir que ha de incluir todos los procesos y actividades que puedan influir en la calidad del producto o servicio de que se trate; es decir, se presta atención a la calidad de las materias primas, del personal, de la organización, servicio post venta, etc., y no sólo los procesos directamente relacionados con la manufactura básica del producto o la prestación del servicio de que se trate.

Como ejemplo más destacado de esta conceptualización cabe citar la explicación que da

Feigenbaum, el primero que introdujo el concepto de *total*, en 1951 en un libro clásico que ha tenido una edición por su 40 aniversario. Feigenbaum, acuñó los términos *total quality control* y *total quality system*, para referirse al control de la calidad en todos los aspectos del ciclo industrial (marketing, diseño/ingeniería, compras, manufactura, inspección del producto, distribución, instalación y servicio post-venta), y a un sistema establecido en toda la empresa para guiar las acciones coordinadas de todos sus elementos en la consecución de la satisfacción del cliente de una forma eficiente. El mismo autor define *total quality management* como aquello que abarca al ciclo completo del producto o servicio, desde su concepción hasta la atención al cliente que adquirió el producto o servicio, pasando por su producción.

No ser *total* significa un ámbito de actuación o atención explícita parcial (por ejemplo sólo a los procesos de producción) tanto si hablamos de control, como de gestión o de sistema.

Si trasladamos las explicaciones de Feigenbaum al concepto de calidad que vimos al principio de este capítulo, cuando empleamos el adjetivo *total* debemos estar incluyendo una atención no sólo a la calidad del producto o servicio en sí, sino también a la calidad de la organización en un sentido amplio, incluyendo relaciones con proveedores y otras organizaciones, así como las relaciones, clima organizacional y satisfacción interna del personal de la propia empresa. ¿Qué ocurre si utilizamos la palabra *total* como adjetivo de *calidad* en vez de como adjetivo de la función (control, gestión, etc.), como de hecho se hace en muchas ocasiones? Pues que, al igual que si la aplicamos a *gestión*, *control* o *sistema*, nos debe quedar claro que lo que estamos indicando es calidad a todos los niveles y de todos los componentes, procesos y relaciones de nuestro negocio, o de la institución, centro o sistema de salud en el que trabajamos.

Lo que debemos evitar es utilizar *total* para significar diferencias en métodos de medición, enfoque o matices prioritarios, en vez de diferencias en el ámbito de aplicación. Así, por ejemplo, a veces encontramos quien dice que las actividades son de *calidad total* porque se miran todas las dimensiones, se emplean mé-

todos tomados de la industria, porque se da mucha importancia a la satisfacción del cliente o porque se tiene en cuenta el componente de costes, etc.; todas estas circunstancias pueden darse perfectamente dentro de un enfoque parcial de la calidad o de su gestión.

Lo lógico es que, cuando hablemos tanto de *gestión total de la calidad* como de *gestión de la calidad total*, lo hagamos para significar que nuestra atención y actividades para mejorar se van a dirigir a todos los ámbitos y niveles, internos y externos, y a todos los servicios o productos, para clientes y de proveedores internos y externos, de nuestra institución.

4. Gestión de la calidad. Componentes esenciales de un sistema de gestión de la calidad

Gestionar la calidad supone la materialización del compromiso con la calidad, traduciéndolo a una estructura organizativa y unas actividades que lo hagan real y operativo. El objetivo de la gestión de la calidad (GC) es la mejora continua, avanzar en la efectividad, eficiencia y satisfacción de todos los protagonistas de la asistencia; un progreso consciente y cuantificable hacia la excelencia.

Sobre esta base, vamos a llamar sistema o programa de gestión de la calidad (GC) al “conjunto de elementos estructurales y de actividades que tienen como fin específico la mejora continua de la calidad”. Este concepto puede encontrarse con diferentes nombres en la literatura sobre calidad, tanto en la industria, como en los servicios de salud. El término globalizador más frecuentemente empleado en la literatura de la industria es el de sistema de calidad. Sin embargo, la terminología empleada en el sector sanitario es diferente y particular. En salud, el término globalizador más frecuentemente empleado es el de *programa* y también, a veces, el de *plan*.

¿Qué terminología adoptar? Si respetamos la tradición sanitaria, le llamaremos programa para enfatizar su carácter finalista y operativo, pero sin renunciar a unos contenidos que incorporan ideas de la industria. Lo que importa en definitiva, se le llame como se le llame, es que podamos identificar una serie de elementos estructurales, organizativos y de activida-

des que se justifican con la existencia previa de una filosofía o política explícita de mejorar continuamente la calidad.

4.1. Los componentes de un programa o sistema de gestión de la calidad

La lista de componentes o elementos que da carácter a un programa de GC puede ser muy larga, según el nivel de detalle al que queramos descender. En sus rasgos organizativos más esenciales o elementales, a un programa de GC lo pueden caracterizar los componentes resumidos en el cuadro 1.3.

El *marco conceptual* incluye el concepto concreto y explícito de calidad (ya hemos visto que hay varias opciones y dimensiones) que se adopte como propio por parte de la organización; la misión, visión y valores que, igualmente explícitos, se pretende compartir con todo

El concepto de programa o sistema de gestión de la calidad incluye la existencia de una política explícita para mejorar la calidad, unos elementos estructurales y organizativos para llevarla a cabo, y unas actividades para hacerla efectiva.

Cuadro 1.3 Componentes de un programa o sistema de gestión de la calidad	
●	Marco conceptual Concepto de calidad Misión, visión, valores Cultura de calidad Calidad total
●	Estructura Recursos Organización
●	Actividades para la mejora continua Ciclos de mejora Monitorización Planificación o diseño
●	Enfoques Recursos Organización
●	Estrategias Liderazgo y responsabilidad (<i>accountability</i>) Sistema de información para la calidad Incorporación de pacientes y población Capacidad organizacional Regulación y estándares Implementación de modelos integrales de atención

el personal de la organización; la búsqueda explícita de una cultura de calidad, y la posible adopción del concepto de calidad total que vimos más arriba.

Una estructura *identificable*, que incluye una comisión o grupo visible que tiene como objetivo promover y preocuparse por las actividades de gestión de la calidad, la integración sinérgica de los diversos niveles de responsabilidad para la calidad dentro de la organización y los recursos dedicados a la función de calidad. La comisión o grupo visible no ha de ser necesariamente quien se encargue de realizar las actividades de evaluación y mejora de la calidad, pero sí ha de promoverlas, apoyarlas y coordinarlas. Entre los rasgos estructurales también se incluyen las normas de funcionamiento, incluyendo reparto claro de responsabilidades, circuitos de información de los datos sobre calidad y mecanismos de elaboración de planes de mejora de la calidad.

Finalmente, las *actividades que se realizan para mejorar la calidad* son tareas identificables y clasificables en razón del fin u objetivo parcial al que sirven, dentro del objetivo común de la mejora de la calidad, y según tres principales puntos de partida en torno a los que puede funcionar el programa: a) la identificación de oportunidades de mejora; b) la elaboración de indicadores y su monitorización, y c) el diseño de nuevos servicios o cambios sustanciales en los existentes. Se trata de tres puntos de partida que se corresponden con tres áreas de ac-

tividades diferenciables dentro del programa: ciclos de mejora, monitorización y diseño de la calidad (cuadro 1.4, figura 1.3).

Cada una de las actividades puede realizarse utilizando métodos y enfoques muy diversos: d) los *enfoques*, interno o externo, desde los cuales se realizan las actividades, y e) los *bloques estratégicos* que se adoptan para avanzar en la mejora de la calidad.

Merece la pena detenernos en los últimos tres componentes y, sobre todo, repasar una visión panorámica e integradora de las tres principales actividades que pueden realizarse para la mejora continua; una de las cuales (los ciclos de mejora) es el objeto principal de este manual.

5. Los bloques estratégicos para la mejora de la calidad

Son varios los ámbitos en los que se puede intervenir para conseguir una mejora importante y sostenida de la calidad. La selección de los ámbitos, o bloques estratégicos, de intervención, representados en la figura 1.3, y de las intervenciones concretas para implementar es la parte esencial de la estrategia de una organización para lograr la mejora de sus servicios.

La consideración de todos los posibles ámbitos, bloques estratégicos o dominios conceptuales para las intervenciones es tanto más importante cuanto más alto sea el nivel del sistema de salud en el que estemos intentan-

Las diversas intervenciones para la mejora de la calidad pueden agruparse en seis grupos o estrategias diferenciadas: liderazgo, sistemas de información, participación de pacientes y población, regulación y estándares, capacidad organizacional, y desarrollo e implementación de modelos de atención.

Cuadro 1.4
Actividades para la mejora continua de la calidad

Grupo de actividades	Punto de partida	Objetivo inmediato
Ciclos de mejora	<ul style="list-style-type: none"> Identificación de un problema de calidad u oportunidad de mejora en algún aspecto de los servicios que se ofrecen. 	<ul style="list-style-type: none"> Solucionar el problema Aprovechar la oportunidad de mejora descubierta
Monitorización	<ul style="list-style-type: none"> Identificación de aspectos relevantes de los servicios que se ofrecen y construcción de indicadores sobre su calidad. Selección de indicadores sobre problemas que hemos sometido a ciclos de mejora. 	<ul style="list-style-type: none"> Identificación de problemas de calidad u oportunidades de mejora
Diseño o planificación	<ul style="list-style-type: none"> Programación de un nuevo servicio para ofrecer. Identificación de necesidades y expectativas de los usuarios. Identificación de parámetros y resultados por conseguir. 	<ul style="list-style-type: none"> Diseñar los procesos de atención para conseguir los resultados deseados predeterminados



Adaptado de: Bengoa R, Key P, Leatherman S, Massoud R, Saturno P: Quality of care. A process for making strategic choices in health systems. WHO, Geneva 2006

Figura 1.3. Los seis grupos de intervenciones para la mejora de la calidad

do actuar; entre otras cosas porque muchas de las posibles intervenciones van a ser posibles y tendrán sentido sólo si se implementan y dirigen desde un nivel relativamente macro (al menos área o distrito).

Los seis componentes estratégicos representados en la figura 1.3. son un marco de referencia útil para poder considerar de una forma integral las diversas alternativas de intervenciones. Lo más común en cualquier sistema u organización es que haga falta algún tipo de intervención en la mayoría de los grupos representados en la figura 1.3. También es frecuente que en la mayoría de los casos se enfatice o se implementen intervenciones en relación con tan sólo uno o dos de los grupos o ámbitos posibles, por ejemplo regulación y estándares, cuando la mejora de la calidad precisa normalmente de un enfoque más integral que contemple intervenciones en la mayoría – si no en todos– los grupos de intervenciones representados en la figura 1.3, cuyo significado es el que se describe a continuación.

5.1. Liderazgo

El liderazgo está situado en el centro del esquema de estrategias en la figura 1.3 porque va ser un componente esencial o, al menos, necesariamente complementario, sea cual sea la intervención que seleccionemos. Es uno de los puntos clave de cualquier estrategia.

El grupo de intervenciones relativas al liderazgo incluye la búsqueda y consolidación del apoyo de los líderes formales y reales (no siempre coincidentes) para las acciones de mejora que queremos implementar. Obviamente los que identifiquemos como líderes para nuestra estrategia van a formar parte del grupo de referencia y seguimiento, porque lo que tenemos que buscar es su compromiso firme con la mejora.

En ocasiones va a ser necesario *construir* la capacidad de liderazgo en las personas que habrían de tenerlo, uno de cuyos componentes va a ser el reforzamiento y reconocimiento de su responsabilidad (*accountability*). Es conve-

Las intervenciones sobre el liderazgo implican la búsqueda y consolidación de los líderes formales y reales, así como un reconocimiento de su responsabilidad.

El sistema de información ha de ser útil para la monitorización de los cambios que queremos lograr, y no tiene por qué basarse siempre en tecnologías sofisticadas.

El incremento de la capacidad organizacional tiene que ver tanto con la atención clínica concreta como con la capacidad de realizar actividades de mejora continua de forma autónoma, así como con el desarrollo de una cultura organizacional que valore la calidad e incluya sistemas de recompensa e incentivos que la favorezcan.

La regulación y estándares son iniciativas normalmente externas, que hay que procurar combinar con los enfoques internos.

La participación de pacientes y población implica el fomento de la participación activa y la puesta en marcha de formas de participación iniciadas o solicitadas por el sistema. Es necesaria para hacer realidad la atención centrada en el paciente.

niente que liderazgo y responsabilidad es van de la mano.

5.2. Sistemas de información

Ya hemos mencionado en el apartado 3 de este capítulo que la reorientación de los subsistemas de información y de formación continua normalmente forma parte de los puntos clave que hay que considerar en la implantación de un programa o sistema de gestión de la calidad en general. Lo mismo ocurre a la hora de diseñar estrategias para mejorar la calidad en temas o problemas de salud concretos. En este caso, debemos asegurarnos un sistema de información que sea útil para la monitorización de los cambios que queremos lograr y el progreso de la estrategia.

El sistema de información requerido no tiene por qué basarse siempre en tecnologías electrónicas sofisticadas. Los cambios deben ser lo más simples posible, con objetivos bien claros (el sistema de información es un medio, no un fin en sí mismo); es posible utilizar, si es preciso, métodos de muestreo y monitorización simplificados, como el LQAS (*Lot Quality Acceptance Sampling*). Es también importante que la información que se genere se ponga a disposición de todas las partes interesadas para reforzar su compromiso, y que se aplique de forma consistente, no parcial, en todos los ámbitos del sistema de salud que estén implicados en la atención al problema objeto de mejora.

5.3. Participación de pacientes población

La aceptabilidad y la atención centrada en el paciente es una de las dimensiones de la calidad. Hay diversas maneras de tener en cuenta al paciente y población para conseguir que esta dimensión sea una realidad. Éstas incluyen formas de participación activa, decididas por el propio paciente, que habremos de fomentar y aprovechar, y formas de participación que podemos solicitarles, las más simples de las cuales son las encuestas sobre su experiencia con el servicio que queremos mejorar.

De forma resumida, las intervenciones en este grupo van a consistir en el fomento de la participación activa, incluida la participación en su propio proceso asistencial, y en la puesta

en marcha de alguna(s) de las formas de participación solicitadas por el sistema o servicios de salud. De nuevo, al igual que decíamos para los sistemas de información, debemos recordar que la participación de pacientes y población debe ser un *medio* para conseguir mejorar. En ocasiones nos encontramos con algunas de estas iniciativas, por ejemplo encuestas, que se implementan como un fin en sí mismas sin que sean utilizadas para mejorar.

5.4. Regulación y estándares

En este grupo se encuentran todas las iniciativas con enfoque externo (explicado más adelante), que abarcan desde los requisitos para el funcionamiento (licencias), los sistemas de acreditación y certificación, y los esquemas de indicadores aplicados de forma general o externa a las instituciones, hasta las guías de práctica clínica y diseño de procesos estandarizados y propuestos para el conjunto del sistema de salud.

Con mucha frecuencia, las iniciativas existentes en los sistemas de salud se encuentran dentro de este grupo, lo que probablemente se deba a su visibilidad y al relativo desconocimiento que aún parece existir sobre otro tipo de intervenciones para mejorar la calidad entre los dirigentes de los sistemas de salud. El principal problema de un énfasis excesivo, o exclusivo, en este tipo de intervenciones de carácter y enfoque típicamente externos es que se olvide el necesario balance que debe existir entre los enfoques externo e interno para lograr la mejora de la calidad.

5.5. Capacidad organizacional

Las intervenciones que se incluyen en este grupo van dirigidas fundamentalmente al desarrollo y gestión de los recursos humanos para facilitar la mejora continua de la calidad, así como a la dotación de la infraestructura y equipamiento necesarios para la provisión de los servicios que queremos mejorar.

Las tres principales áreas de intervención son:

- El desarrollo de recursos humanos con los conocimientos y habilidades necesarios para la prestación de los servicios requeri-

dos, y con el equipamiento necesario para su trabajo.

- El desarrollo de sistemas y capacidad metodológica para la práctica de la mejora continua de la calidad (ciclos de mejora, monitorización, planificación de la calidad), con énfasis en el compromiso interno y la autoevaluación.
- El desarrollo de una cultura organizacional que valora la calidad, incluyendo, en su caso, sistemas de recompensa e incentivos que la favorezcan.

En este grupo se incluye la atención que siempre hay que prestar a la reorientación del subsistema de formación continuada, tal como vimos en el apartado 3 de este capítulo.

5.6. Desarrollo e implementación de modelos de atención

Los modelos de atención son la forma más completa y compleja de la planificación o diseño de la calidad. Se trata de diseñar e implementar modelos de atención integrales e innovadores que incluyan toda la evidencia existente en torno al problema de salud que queremos mejorar. Uno de los más conocidos es el modelo de atención a enfermedades crónicas* que requiere de una coordinación expresa y precisa de todos los niveles de atención, así como de una participación activa y responsable de los pacientes. Otros ejemplos destacados se han desarrollado en el ámbito de la salud materno-infantil, partiendo del llamado continuum de cuidados propuesto por WHO, UNICEF y UNFPA, que comienza con la salud reproductiva y termina con la atención a la adolescencia, pasando por la atención prenatal, perinatal e infantil.

Los modelos de atención pueden enfocarse en una parte concreta del continuum de cuidados o de la atención al problema de salud considerado, pero siempre han de incluir toda la evidencia existente para todos los niveles de atención implicados.

6. Las actividades para la mejora de la calidad

Las tres grandes iniciativas o actividades que dan entrada a la dinámica de la mejora continua para servicios específicos (ciclos de mejora, monitorización y planificación o diseño de la calidad) se diferencian por el punto de partida que las inicia y el objetivo inmediato que persiguen, tal como se resume en el cuadro 1.4 y describimos a continuación.

6.1. Ciclos de mejora

La evaluación, el ciclo de mejora, ciclo evaluativo o ciclo de garantía de calidad de la literatura tradicional sobre calidad en los servicios de salud, comienza con la identificación de una oportunidad de mejora en algún aspecto de los servicios que se están ofreciendo (el “problema” de calidad en la terminología tradicional). Este problema puede ser de tipo asistencial (calidad de los programas, diagnósticos, tratamientos, controles de salud, etc.), organizativo (descoordinaciones, tiempos de espera, circuitos anómalos, etc.) o de cualquier otro tipo (satisfacción, comunicación, información, etc.). El objetivo de las actividades que se generan a partir de que el tema a estudio (*oportunidad de mejora o problema de calidad*) se ha identificado y definido es elevar el nivel de calidad del aspecto que se evalúa a los niveles máximos que permite el contexto en el que se prestan los servicios. Es decir: *aprovechar* la oportunidad de mejora o *solucionar* el problema identificado.

A veces, sólo con el análisis del problema y la consideración de los criterios o requisitos de calidad que queremos, resulta evidente lo que hay que hacer para mejorar y se puede diseñar directamente la intervención oportuna. Otras veces, sobre todo cuando el problema puede limitarse a un solo requisito de calidad (tiempos de espera elevados, insatisfacción, fallos organizativos puntuales y repetidos, etc.) de difícil comprensión y sin relación sabida con criterios de calidad concretos, lo oportuno es hacer un estudio de causas o comprobación de hipótesis que expliquen el incumplimiento del criterio o requisito de calidad en estudio. Estas dos eventualidades están marcadas en el gráfico más completo (figura 0) como vías de

A pesar de su inutilidad como actividad aislada, la monitorización resulta muy atractiva y a veces es la única actividad sobre calidad que se lleva a cabo.

* http://www.improvingchroniccare.org/index.php?p=The_Chronic_Care_Model&s=2 (último acceso en Septiembre 2015)

Los ciclos de mejora son una actividad imprescindible en los programas o sistemas de gestión de la calidad.

acción contingentes o alternativas y completan las actividades que se pueden relacionar directamente con la evaluación de la calidad, el clásico *ciclo evaluativo* o *ciclo de garantía de calidad*, en el que se evalúa la presencia o ausencia de criterios o requisitos de calidad que se sabe de antemano que son importantes (por ejemplo, todos los requisitos de una actuación clínica adecuada a un tipo de paciente o problema de salud concreto).

El ciclo de mejora es probablemente lo más conocido y fácil de entender, pero también es el grupo de actividades central y crucial de cualquier programa de gestión de la calidad y el que más necesita de un enfoque interno, de autocompromiso. Sin embargo, no es suficiente si queremos tener un programa completo, aunque podamos encontrar programas que se basen exclusivamente en ciclos de mejora.

6.2. Monitorización de la calidad

Entendemos la monitorización como una medición periódica, planificada y rutinaria de una serie de aspectos-resumen o indicadores de la calidad. Su objetivo es asegurarnos que nos mantenemos a unos niveles previamente fijados como aceptables. En la práctica, para la monitorización se parte de una selección de aspectos o servicios para los que se elaboran los correspondientes indicadores, que han de ir acompañados de su correspondiente esquema de medición o plan de monitorización, en el que hay que especificar, entre otras muchas cosas, el método y la periodicidad de las mediciones. Estos indicadores pueden derivarse de ciclos de mejora realizados o ser de entrada contruidos y seleccionados para monitorizar aspectos que consideramos relevantes en los servicios que ofrece el centro. El resultado o fin principal de la monitorización es identificar oportunidades de mejora, o problemas de calidad que merecen una atención adicional más profunda y detallada.

La monitorización es igualmente útil en los programas o enfoques internos y externos, como después veremos, no como los ciclos de mejora que tienen su máxima aplicabilidad en los programas internos. A veces, junto con la acreditación que es otra forma de monitorizar si se hace de forma periódica, la monitoriza-

ción es la única actividad desarrollada en los programas externos (de iniciativa externa al centro). Piénsese, por ejemplo, en los listados de indicadores o normas que se manejan para evaluar desde los niveles centrales la atención que se presta en los diversos centros sanitarios.

6.3. El diseño o planificación de la calidad

Si los ciclos de mejora *curan* problemas que ya existen, y la monitorización tiene como fin detectar situaciones problemáticas o mejorables, diseñar la calidad es lo mismo que prevenir la aparición de problemas de calidad e implica, en la terminología de la industria, diseñar sistemas y procesos de manera que la calidad sea inevitable; es decir, facilitar que las cosas tengan que hacerse siempre bien, de manera que el resultado sea el que precisamente hemos fijado como esperado.

El diseño de la calidad es el área de actividades que ha aparecido más tarde en los programas de gestión de la calidad, pero también a la que más importancia se le está dando y la más compleja de dominar. Su práctica parte de definir qué queremos conseguir, en términos de expectativas y necesidades por satisfacer, resultados o parámetros del servicio por diseñar y, a partir de ello, decidir qué hacer y cómo hacerlo para lograr el resultado esperado cuando tengamos ocasión de ofrecer el servicio que hemos diseñado.

Lo común dentro de las actividades de diseño de la calidad son los diseños de intervención subsiguientes al análisis del problema o la discusión de los datos en un ciclo de mejora; el diseño de intervenciones es el punto en el que se unen evaluación y diseño de la calidad (figura 0). Por otra parte, en los servicios de salud hay una cierta tradición en diseñar la calidad de la atención clínica: la construcción de protocolos o guías de práctica. En la actualidad, el enfoque de diseño, con diversos nombres y acepciones (uno de los más populares, *reingeniería*) es el que se propone como fundamental para los programas de GC; incluso se llega a aconsejar empezar por ahí. Sin embargo, diseño solo tampoco es suficiente. Por muy perfecto que pensemos que sea, siempre hay que contrastarlo con la realidad, es decir, evaluar los criterios o indicadores que nos digan si se

La monitorización de la calidad es una medición periódica y planificada de una serie de indicadores, con el fin de controlar que están a los niveles deseables. Sirve para identificar problemas de calidad u oportunidades de mejora y controlar cambios.

Con el diseño de la calidad se intenta que las cosas se hagan siempre bien y a la primera según unos parámetros prefijados, previniendo la aparición de problemas de calidad.

está implantando como estaba previsto y si funciona o no.

6.4. Otros modelos para representar las actividades de un programa de GC: ciclo de garantía de calidad, trilogía de Juran, ciclo de Deming

El esquema de actividades que hemos propuesto y explicado tiene sus raíces en el ciclo de garantía de calidad ampliamente asumido en el pasado en el ámbito del sector salud, y que hemos ido modificando a medida que la conceptualización y la práctica de los programas de gestión de calidad han necesitado de modelos más completos. El clásico ciclo de garantía de calidad o ciclo evaluativo sólo contempla explícitamente las actividades propias de lo que hemos llamado ciclos de mejora, y aun así con algunas dificultades de entendimiento en su aplicación cuando el problema identificado no es de tipo clínico. Por tanto, no puede evaluarse la calidad directamente con base en criterios de buena práctica previamente conocidos y establecidos, cuya falta de cumplimiento sería la causa directa del problema de calidad.

El esquema de actividades completo, compuesto por ciclos de mejora, monitorización y diseño de la calidad, que hemos ido desarrollando por la necesidad de modelar adecuadamente la realidad práctica y la visión de los programas que íbamos construyendo en centros sanitarios, se asemeja mucho a la esencia de otros esquemas utilizados en el ámbito de la industria. El más parecido y conocido es la llamada *Trilogía de Juran*. Los términos que se incluyen en el modelo de Juran, con alguna variación –lo que aquí hemos llamado diseño, monitorización y ciclos de mejora– son actividades de planificación, control y mejora, aunque este autor no ha descendido, como en nuestro caso, al detalle de los pasos que hay que incluir en cada grupo de actividades.

Otros esquemas y terminologías sobre las actividades que comprende un programa de gestión de la calidad (o cualquiera de los diversos nombres que pueda recibir: garantía de calidad, gestión de la calidad total, mejora continua de la calidad, etc.) se cuentan por decenas. Dentro de la terminología o modelo

conocido como *gestión de la calidad total* (*total quality management*), han surgido multitud de esquemas. La mayoría de estos esquemas representan sólo uno de los tres grupos de actividades (normalmente, los ciclos de mejora), confunden innecesariamente los tres, o son adaptaciones de otros esquemas con gran predicamento en la industria que se llegan incluso a desvirtuar. El ejemplo más paradigmático e ilustrativo de esto probablemente son las múltiples adaptaciones del ciclo de Shewart, que Deming popularizó en Japón y que actualmente se conoce como PDCA (*plan-do-check-act*), PDSA (*plan-do-study-act*) o ciclo de Deming. El ciclo de Shewart fue mostrado por Deming como un esquema para la mejora de la calidad con base fundamentalmente en el diseño de innovaciones (*plan*, planificar), que luego hay que ensayar a pequeña escala (*do*, hacer) y medir sus efectos (*check*, comprobar o estudiar), para después incorporar los rasgos positivos a la producción rutinaria de ese producto o servicio (*act*, actuar). La inevitable interconexión entre los tres grupos de actividades ha ido produciendo adaptaciones varias que explicitan más pasos de las otras dos (monitorización y solución de problemas), en detrimento de la idea inicial de ser un ciclo que se inicia como actividad de diseño (innovación). En estado *puro* o en una de sus múltiples adaptaciones, el PDCA o PDSA (como también se le llama, al sustituir el original y más vulgar *check* por el más académico *study*) es probablemente el modelo más citado, adaptado y referenciado en la actualidad.

6.5. Para qué sirve un sistema de GC

La visión de lo que entendemos por un sistema o programa de GC no estaría completa sin una idea clara de lo que se puede conseguir en la práctica; es decir, de cuáles son sus resultados y su justificación.

Podemos afirmar, sin temor a equivocarnos, que el compromiso explícito, estructurado y funcional con la calidad sirve para aumentar la efectividad, la eficiencia (a través sobre todo de una reducción drástica de los costes de la calidad deficiente) y la satisfacción de proveedores y usuarios, y además nos pone en el camino hacia la excelencia. Si se juega bien, es un

La trilogía de Juran (planificar, controlar y mejorar la calidad) se corresponde con lo que hemos llamado diseño, monitorización y ciclos de mejora.

Hay multitud de otros esquemas, de los cuales el más conocido es el PDCA o *Ciclo de Deming*, profusamente adaptado e interpretado de maneras muy diversas.

Los programas o sistemas de gestión de la calidad son el camino hacia la excelencia, punto culminante de la autoestima.

Los programas o sistemas de Gestión de la Calidad aumentan la efectividad, la eficiencia y la satisfacción de proveedores y usuarios

juego sin perdedores: los gestores tienen ante sí una formidable herramienta para orientar su gestión; los profesionales de la salud trabajan mejor y, dado que responden de forma reconocida y responsable a sus necesidades éticas y científicas, se sienten satisfechos; la población atendida sólo puede esperar beneficios en su salud y satisfacción cuando mejora la calidad de la atención sanitaria que recibe. La evidencia empírica es abundante en la tradición de los programas de calidad tanto en la industria como en los servicios de salud.

En el campo industrial, la representación más paradigmática de los beneficios que puede aportar la mejora de la calidad es la llamada “reacción en cadena” de Deming: mejor calidad → más productividad → disminución de costes → disminución de precios → aumento de cuota de mercado → crecimiento de la empresa y de los beneficios. Sin embargo, el énfasis se pone fundamentalmente en los costes y los beneficios económicos, lo cual es comprensible porque es el lenguaje que mejor entienden los gestores y, al parecer, esta literatura va dirigida sobre todo a este tipo de lectores; pero hay más beneficios que los puramente económicos. Por ejemplo –y continuando en el amplio campo de la industria–, un resumen de los beneficios de la GC medidos en 22 compañías finalistas del Baldrige National Quality Award de Estados Unidos incluye, además de una mayor productividad y mayores beneficios de mercado, mejor clima laboral y mayor satisfacción los sus clientes. En el sector salud la situación es la misma. Las publicaciones que describen los resultados de las iniciativas de GC documentan mejoras muy concretas, cuantificadas e importantes en la efectividad, eficiencia y satisfacción de personal y usuarios.

Hasta aquí, todo lo mencionado consiste en efectos y beneficios que tenemos al alcance de la mano y que podemos ver materializados en el corto y medio plazo, pero hay también otros efectos más intangibles y que van cobrando sentido a más largo plazo: los programas de GC son el camino hacia la excelencia, un objetivo final, cambiante, e inalcanzable por los propios cambios a que está sujeta la visión de la excelencia, pero motor último del amor propio. Fernando Savater lo ha resumido de forma muy acertada:

El anhelo de excelencia y perfección, culminación del arte de vivir ético, es el producto más exquisito del amor propio adecuadamente considerado. Quien no desea ser excelente ni perfecto, quien crea que no se merece tanto o no se atreve a proponerse tanto, es que desde luego no se ama lo suficiente a sí mismo; o tiene una idea de la excelencia y perfección puramente ajena, pervertida, esclavizadora, incompatible con las urgencias inaplazables de su yo.

Los programas de GC nos ayudan a ser excelentes, es decir, exquisitamente egoístas, y a disfrutar con ello.

6.6. Razones para no tener un programa de GC

Si tenemos en cuenta el compromiso ético de los profesionales de la salud con los pacientes a los que atienden, el rechazo a los programas de GC sólo podría estar basado en que la alternativa de no tenerlos fuese más beneficiosa para la población atendida o para los intereses de los profesionales. Igual argumentación podría utilizarse, tanto para los que proveen atención directa al paciente, como para el nivel gestor y otros proveedores de clientes internos.

Por otra parte, la evidencia empírica sobre los programas de GC es constante y consistentemente la misma: los beneficios son tangibles y manifiestos en comparación con su inexistencia. ¿Cómo se explica entonces que lo más frecuente sea que no exista programa de GC? Los posibles argumentos para responder pueden resumirse en tres: ignorancia, inercia e incompetencia.

- *Ignorancia.* Está claro que la existencia de programas de GC no es exigible si no se sabe previamente en qué consisten y cómo ponerlos en marcha. La calidad no es aún un tema sobre el que tengan conocimiento y formación todos los profesionales de la salud ni todos los gestores.
- *Inercia.* Saber no es suficiente; también hay que querer y poder superar los condicionantes adversos que se imponen reforzados con el tremendo poder de la inercia, fácilmente asociada a la indecisión. Se puede argumentar que estos condicionantes adversos llegan a veces a hacer imposible el programa de GC. En realidad, lo que

La no existencia del sistema de GC puede justificarse por ignorancia, inercia, incompetencia o una mezcla de las tres. Para ponerlo en marcha hay que saber, querer y poder.

puede ocurrir es que lo hagan más difícil, no imposible. Es en estas situaciones en las que es más evidente la importancia de la capacidad creativa y el ejercicio del liderazgo profesional para ir abriendo brecha. Se trata de dar el paso de tener un compromiso explícito y operativo con la calidad a partir del compromiso implícito que todo profesional tiene.

- *Incompetencia.* Cuando se sabe en qué consisten los programas de GC y existe la posibilidad cierta y evidente de ir venciendo la inercia, el no ponerlos en marcha es ejercer la profesión de forma incompleta. Éste es un síntoma de cierta incompetencia, sobre todo en el nivel gestor.

7. Programas internos y externos en la estrategia para la mejora de la calidad

Una característica importante de los programas e iniciativas para la mejora de la calidad es el énfasis específico o la combinación de iniciativas externas e internas. Los enfoques interno y externo se distinguen en función de que las actividades partan de los implicados directamente en aquello que se evalúa (autoevaluación, enfoque interno) o se pretenda evaluar las actuaciones ajenas (enfoque externo).

En el sector salud, algunos autores distinguen una tercera categoría que podríamos llamar enfoque externo en programas internos, en referencia a las evaluaciones que se pueden hacer de unos profesionales sobre otros dentro de un mismo centro de salud u hospital de acuerdo con una línea jerárquica o de otro tipo (por ejemplo, de tutores sobre sus residentes, de gerentes sobre los servicios, de coordinadores sobre coordinados, etc.). Estas evaluaciones son internas en relación con la institución, pero no con el profesional evaluado.

En el *enfoque interno*, los que son evaluados quieren la evaluación, están implicados en el establecimiento de los criterios e indicadores que van a ser utilizados y participan, de forma activa o delegada, en la totalidad de actividades de los ciclos de mejora, monitorización y diseño. La utilidad de las iniciativas internas depende directamente del compromiso de los propios evaluados de ser rigurosos con su evaluación y con la mejora de la calidad. Una

de las principales ventajas de este enfoque es que reconoce el grado de responsabilidad y autonomía que los profesionales de la salud tienen indudablemente en la prestación de los servicios; además de que es de manera interna como mejor se entiende el contexto en que se da la asistencia y los medios disponibles para mejorarla. Como consecuencia, los programas internos se adaptan a las condiciones locales; se enfatiza la profesionalidad y se canaliza el deseo de la mayoría de hacer las cosas de la mejor manera posible; además, es lógico que el ejercicio interno del compromiso con la calidad facilite la sinergia con las iniciativas externas para el mismo fin.

Entre los inconvenientes, al principio está el de tener que buscar el tiempo y esfuerzo necesarios para autoevaluarse y mejorar; una vez superado éste pueden aparecer otros, como, por ejemplo, determinados conflictos de autoridad, desarrollo de métodos demasiado idiosincrásicos con poca base científica, autocomplacencia por la falta de una referencia externa para compararse, o falta de intervenciones y seguimiento sobre determinados problemas internos que pueden aparecer.

En el *enfoque externo*, la iniciativa en sí para evaluar y las normas de la evaluación o monitorización son establecidas por personas distintas de los proveedores que van a ser juzgados.

Su principal ventaja es que, al menos en teoría, sirven para un propósito socialmente más amplio: el que corresponde a todo el sistema de salud en cualquiera de sus niveles. Adicionalmente, los programas externos permiten las comparaciones entre centros semejantes, pueden llamar la atención sobre problemas que de otra forma pasarían inadvertidos y hacen mucho más difícil la autocomplacencia o la ocultación de problemas. Sin embargo, uno de sus principales inconvenientes es que revisan características de fiscalización o inspección, con el consiguiente rechazo de los inspeccionados, en vez de ayudar a la evaluación interna, a la identificación de problemas y, en definitiva, al proceso de mejora de la calidad. Los programas externos precisan también de una considerable infraestructura organizativa y, sobre todo, de utilizar normas, criterios e indicadores suficientemente fiables y válidos para

su propósito. Los programas externos pueden utilizar estándares artificiales o irrelevantes, desconociendo o ignorando las circunstancias de la asistencia diaria y, lo que es peor, los medios con los que se puede mejorar la asistencia en un centro concreto.

Los enfoques interno y externo se complementan; ambos son necesarios para un acercamiento global de sistema de salud a la mejora de la calidad, y ninguno es por sí solo una respuesta suficiente desde el punto de vista de la población en su conjunto. Sin embargo, aun insistiendo en la complementariedad de los dos tipos de enfoque, lo que tenemos al alcance en cualquiera de los niveles que nos encontremos en el sistema de salud son las actividades de evaluación y autocorrección desde la perspectiva interna. Pero aún hay más razones para traer a primer término el enfoque interno. Ya hemos comentado cómo el entender y practicar la gestión de la calidad como algo inherente al funcionamiento diario de los respectivos centros y servicios favorece la superación de resistencias frente a propuestas de sistema que pueden ser percibidas como fiscalización del exterior, y no como una necesidad de optimizar profesional y socialmente los servicios de salud. Adicionalmente, los análisis realizados de los muchos años de experiencia con programas de todo tipo en el multisistema de Estados Unidos concluyen en lo imprescindible que resultan los programas internos para lograr la mejora de la calidad, poniendo en primer término el interés y la responsabilidad de los profesionales que la protagonizan. En esencia, los programas externos pueden apoyar u obligar a que existan programas internos, pero la mejora de la calidad exige que la gestión de la calidad sea asumida como un esfuerzo interno. No puede parecer lógica ninguna estrategia de mejora de la calidad que descansa principalmente ni, menos, exclusivamente en acercamientos externos. Esto no quiere decir que los programas externos no sean útiles o necesarios: la situación ideal es de sinergismo entre programas internos y externos, donde los programas externos apoyan a los internos y ofrecen elementos de comparación entre instituciones de características semejantes. Lo que sí es cuestionable es intentar que se empiecen las actividades de mejora de la calidad en las

instituciones de salud sobre la base de un enfoque exclusivamente externo. Como resume Palmer: “desde fuera se puede medir, desde dentro se puede medir y mejorar”.

8. Responsabilidades en la mejora de la calidad

La calidad en los servicios de salud tiene tres grupos de protagonistas directos (personal de gestión, personal que provee directamente los servicios y población atendida), y los tres tienen algo que decir y hacer para mejorar la calidad. El problema no es, sin embargo, identificar quiénes son los protagonistas sino tener claro cómo ejercer este protagonismo. Para ello, en primer lugar, hay que tener presentes los dos tipos de enfoques y programas complementarios: externo e interno, y después hay que revisar cuáles de ellos están al alcance de cada uno, cuáles son sus principales actividades y cómo se justifica su realización para conseguir el objetivo final de la mejora continua de la calidad.

8.1. El papel de los gestores

Una de las principales características de los más recientes enfoques de los programas de GC es el reconocimiento de la máxima responsabilidad que tiene el personal gestor para que el programa exista y funcione. En los libros y manuales sobre implementación de programas de GC con los métodos tomados de la industria se afirma, de forma totalmente inequívoca, que sin el compromiso activo de los directivos no hay programa posible.

Los necesarios programas externos, tanto los de nivel de sistema como los de niveles intermedios, son una responsabilidad ineludible del nivel gestor. La forma en que estén diseñados e implementados y la medida en que contribuyan o no a la mejora de la calidad dependen ampliamente de la capacidad de gestión de sus impulsores. El nivel gestor es también responsable de impulsar y mantener los programas internos de los centros a su cargo; pero para ello, todos los autores subrayan la necesidad de tener en cuenta las peculiaridades de la actividad y el personal que en la práctica ha de llevar adelante las actividades de GC y, en defi-

La gestión de la calidad tiene tres protagonistas: gestores, personal de salud y población que recibe los servicios.

nitiva, la mejora de la calidad. La participación del personal sanitario es uno de los factores clave para el éxito del programa.

8.2. El papel del personal de salud

El personal encargado de proveer directamente la asistencia sanitaria es la base insustituible de los programas internos y de todo el sistema. Sin su compromiso y participación, la mejora de la calidad no es posible aunque exista oficialmente un programa de GC. En uno de los primeros libros sobre las experiencias del enfoque industrial aplicado a la salud, editado entre otros por Donald Berwick, se reconoce taxativamente. Por otra parte, el grado de autonomía en las decisiones a la hora de la prestación de servicios que tiene el personal de salud probablemente sea de los más elevados que puedan encontrarse en cualquier empresa. En muchos de los aspectos, un solo médico, enfermera, fisioterapeuta o un mínimo equipo son, en gran parte, gestores de sí mismos, de su tiempo y de sus acciones, y las decisiones que tomen tienen una gran trascendencia para la calidad del servicio que ofrecen. Adquieren así, a su nivel, las mismas responsabilidades que puedan ser atribuidas a los gestores de niveles superiores de la organización sanitaria, aunque dentro de su particular ámbito de influencia.

Desde otra perspectiva, la participación del personal de salud en los programas internos de los centros asistenciales en su autoevaluación para asegurar y mejorar la calidad de la atención que ofrecen puede justificarse por razones eminentemente prácticas: no quedarse al margen de algo que se va imponiendo hasta hacerse inevitable. Sin embargo, su justificación más profunda son las razones de ética profesional; la evaluación y mejora de la calidad es un deber que tienen consigo mismos y con sus pacientes. El rechazo de los programas de GC sólo sería justificable si la alternativa de no participar se mostrase más beneficiosa para los intereses de los pacientes –de los que son ineludiblemente responsables– y del propio personal de salud; éste no parece ser el caso a la vista de los abundantes problemas de atención, muchos de ellos previamente inadvertidos, que afloran cada vez que se evalúa la calidad de los servicios de

salud. Adicionalmente, en el sector salud la influencia de los líderes profesionales puede ser mucho mayor que la de los gestores a la hora de introducir actividades de GC. Este hecho, comprobado empíricamente en más de una ocasión, confiere una responsabilidad adicional a los buenos profesionales en la práctica pionera de las actividades de GC.

8.3. El papel de los usuarios

De una forma esquemática, podemos clasificar las principales formas en que el usuario puede contribuir a la GC en dos grandes grupos metodológicos: activa (por iniciativa propia) o solicitada por el sistema de salud. Metodológicamente, la distinción es importante pero, al igual que ocurre con los enfoques interno y externo, no deben considerarse como excluyentes. Ambas son útiles, ambas son teóricamente necesarias y, por tanto, el grado de desarrollo de ninguna de ellas debería oscurecer a la otra, ni impedir que se avance en el conocimiento de métodos y mecanismos de optimización de cada una de las formas de participación posible. No puede decirse, sin embargo, que se haya demostrado un nivel de utilidad en GC parecido para todas ellas, ni que la metodología de su puesta en marcha y aprovechamiento efectivo esté igualmente desarrollada. En la actualidad, aun dentro de las muchas dificultades y puntos oscuros que todavía existen, se sabe mucho más de los métodos y aprovechamiento rutinario de la participación solicitada (por ejemplo, a través de encuestas y documentos como el consentimiento informado y las voluntades previas) que de las diversas formas de participación activa, entre las que destacan las quejas y reclamaciones y la participación activa en el propio proceso asistencial.

9. Estrategias de implantación de actividades. El ciclo de mejora

Como vimos en el apartado 6 de este capítulo, podemos identificar tres grandes grupos de actividades en los programas de GC, que responden a otros tantos puntos de partida: diseño de la calidad, monitorización y ciclos de mejora. Todos ellos terminan estando interrelaciona-

El rechazo a los programas o sistemas de GC sólo se justificaría si el no tenerlos fuese mejor para los pacientes y los propios profesionales.

El grado de autonomía en las decisiones que afectan a la calidad, y la influencia del liderazgo profesional en la implementación de actividades de GC son dos razones adicionales para subrayar el importante papel de los profesionales.

El usuario de los servicios puede participar por decisión propia o porque se le solicite. Ambas formas son útiles para la GC.

dos, pero en teoría es posible comenzar nuestras actividades por cualquiera de los tres. Sin embargo, lo que vamos a proponer es la realización de un ciclo de mejora, ¿por qué? Hay tres razones: la primera es que, dentro de su posible complejidad, es la actividad más fácil de realizar y comprender; la segunda es que, como actividad de aprendizaje, es muy completa y básica, de modo que abona el terreno para comprender y practicar mejor las otras dos, y la tercera es que posibilita la obtención de resultados a un relativo corto plazo, de modo que motiva a continuar.

Schein, uno de los grandes nombres del desarrollo organizacional, resume en siete puntos sus recomendaciones estratégicas para introducir cambios; destaquemos los dos que se refieren a actividades concretas: “comenzar como si fuese un experimento” y “empezar con pasos poco ambiciosos y adoptar una estrategia de implantación progresiva”. Aplicado a nuestro tema, comenzar por actividades de diseño, aunque es tremendamente atractivo e incluso hay una tendencia actual a proponerlo como primer paso, no sería una buena opción porque reviste en general mayor complejidad que los otros dos. También es muy frecuente que se propugne comenzar por elaborar un listado de indicadores y monitorizar. Sin embargo, optar por la realización de un ciclo de mejora es un comienzo más simple, fácil de enfocar y que a la vez resulta enriquecedor y didáctico. Sin embargo, no es descartable ni excluyente comenzar identificando problemas a través de la medición de indicadores bien seleccionados, o bien diseñando o rediseñando un servicio especialmente importante. Todo depende del grado de madurez, compromiso y recursos iniciales para la GC: han de ser mayores para comenzar con monitorización, y aún más para comenzar con diseño. En relación con la monitorización, ha de quedar claro, además, que monitorizar no es, en sí mismo, mejorar. Monitorizar es asegurarse de que nos mantenemos a unos niveles de calidad aceptables y preestablecidos, o hacer despistaje de existencia de problemas, mientras que evaluar en un ciclo de mejora es buscar las causas y lograr la solución de esos problemas. Los ciclos de mejora son la esencia imprescindible de cualquier programa de GC, tienen sentido por sí mismos. La moni-

torización es necesariamente una actividad complementaria que, aunque resulta de gran relevancia en los programas estables, es perfectamente prescindible al inicio de los mismos.

Una secuencia posible, e incluso deseable, en la implantación de actividades de GC es que, una vez conocidos y practicados los métodos y herramientas útiles para los ciclos de mejora, se sintiera la necesidad de conocer también las de monitorización y diseño, y de liderar la implantación de la GC como actividad plenamente asumida en el entorno de trabajo de cada uno. En este manual vamos a centrarnos en lo que supone la realización de un ciclo de mejora, cuyos pasos se describen brevemente a continuación. En cada paso nos auxiliaremos de determinados métodos y herramientas, los cuales detallaremos en cada capítulo de este manual, de acuerdo con el esquema representado en la figura 0, al principio de este capítulo.

9.1. Pasos de un ciclo de mejora

Paso 1. Identificación y priorización de la oportunidad de mejora. El ciclo de mejora comienza con la identificación de una oportunidad de mejora o problema de calidad. A partir de ahí se generan actividades cuyo objetivo es solucionar el problema identificado o, lo que es lo mismo, aprovechar la oportunidad de mejorar. De los muchos métodos que se pueden utilizar, hay que conocer a fondo por lo menos unos cuantos de tipo cualitativo, de los que destacamos la *lluvia de ideas*, la *técnica de grupo nominal* y las *matrices decisionales*, aparte de otros que precisan de datos y una determinada infraestructura para realizarlos. En este paso debemos elegir cuál es el servicio o el aspecto del mismo que queremos mejorar.

Paso 2. Análisis del problema de calidad. Es crucial para decidir qué hacer a continuación. Para el análisis de problemas de calidad son de utilidad diversos métodos como el *diagrama de causa y efecto*, *flujograma*, *histograma*, *estratificación* y *diagrama de Pareto* y *gráficos de control*, utilizables con y sin datos previos sobre el problema, según el método. De este paso se derivan tres posi-

Proponemos comenzar la dinámica de la mejora continua con la realización de un ciclo de mejora por su relativa facilidad de comprensión, por su valor didáctico para la GC y por su efecto motivador si se realiza con éxito.

Los ciclos de mejora se realizan siguiendo una serie de pasos concretos, que se benefician de métodos y herramientas relativamente específicos que serán objeto de revisión en los diversos capítulos de este manual.

bles cursos de acción: diseño de intervención, diseño de estudio de nivel de calidad y diseño de estudio de causas hipotéticas. En cualquier caso se necesita previamente tener claro cuáles son los criterios o indicadores que nos midan la existencia o no de calidad.

Paso 3. Construcción de criterios para evaluar calidad. En este paso deben identificarse criterios válidos y fiables para medir la calidad del servicio o aspecto del servicio en estudio.

Junto a las indicaciones formales para valorar la validez, hay que manejar un esquema para medir la fiabilidad, antes de utilizar los criterios o indicadores de forma generalizada.

Paso 4. Diseño de estudio de nivel de calidad o de causas hipotéticas. En la realización de ciclos de mejora hay que detallar los componentes para el estudio de nivel de calidad, y los de causas hipotéticas que puedan realizarse incluso después de haber realizado un estudio de nivel de calidad. En este paso hay que utilizar unos conocimientos básicos de estadística para beneficiarse de las ventajas de evaluar con muestras.

Paso 5. Análisis y presentación de los datos de la evaluación. Tras recoger los datos de acuerdo con el diseño realizado, hay que buscar la forma más eficiente de presentarlos y analizarlos. Si hemos evaluado una muestra, daremos los datos estimados con intervalos de confianza, y veremos la forma de

realizar un análisis gráfico con diagrama de Pareto y otras representaciones visuales, de manera que nos resulte más evidente sobre qué incidir para mejorar.

Paso 6. Diseño de intervenciones para mejorar. La discusión de los datos de la evaluación puede sugerir por sí sola qué hacer para mejorar; sin embargo, es conveniente conocer una serie de herramientas que pueden ser útiles para decisiones grupales sobre diseño de planes de acción para la mejora. Entre estas herramientas, está el *análisis de campos de fuerza*, el *diagrama de afinidades* y la *votación múltiple*, que deben ser revisados para ayudar en la realización de este paso.

Paso 7. Implementación de la intervención diseñada. En esencia, es el cumplimiento del plan de acción resultante del paso 6.

Paso 8. Reevaluación y documentación de la mejora conseguida. Es el paso que cierra el ciclo de mejora si hemos tenido éxito con la intervención diseñada e implementada. Para documentar la mejora, hemos de efectuar la comprobación estadística de que la diferencia es significativa, si hemos utilizado muestreo, y una forma sencilla y muy completa de visualizarlo gráficamente como, por ejemplo, un diagrama de Pareto antes-después. Tras la reevaluación, podemos decidir monitorizar, reanalizar el problema –si no hemos conseguido mejorar– o emprender otros ciclos de mejora sobre un aspecto o servicio diferente.

BIBLIOGRAFÍA

- Bengoa R, Kavar R, Key P, Leatherman S, Masson R, Saturno P. Quality of care. A process for making strategic choices in Health Systems. Geneva: WHO, 2006. Cap 2.
- Berwick DM, A Blanton Godfrey AB, Roessner J. Curing Health Care. New Strategie for Quality Improvement. San Francisco: Jossey Bass, 1990.
- Deming WE. Calidad, productividad y competitividad. La salida de la crisis. Madrid: Díaz de Santos, 1989. Cap. 6.
- Donabedian A. Explorations in quality assessment and monitoring. Vol. I. The definition of quality and approaches to its assessment. Ann Arbor: Health Administration Press, 1980.
- Feigenbaum AJ. Total Quality Control, 3ª ed. (edición 40 aniversario). New York: McGraw Hill, 1991.
- Institute of Medicine. Crossing the Quality Chasm. A New Health System for the 21st Century. Washington: National Academy Press.
- Ishikawa K. What is quality control? The Japanese way. Englewood Cliffs: Prentice-Hall, 1985. Cap. 3.
- Juran JM, Gryna FM, Bingham RS. Manual de Control de la Calidad, 2ª ed. Reverté, Barcelona, 1990. Cap. 2 y 3.
- Leebow W, Ersoz CJ. The health care manager's guide to continuous quality improvement. Chicago: AHA, 1991, Cap. 11.
- Lohr KN (ed.). Medicare: A Strategy for Quality Assurance. Washington DC: National Academy Press, 1990. Vol. I, Cap. 9.
- Office of Technology Assessment (OTA). Assessing the Efficacy and Safety of Medical Technologies. Washington DC: US Government Printing Office, 1978.
- Palmer RH. Evaluación de la asistencia ambulatoria. Principios y práctica. Madrid: Ministerio de Sanidad y Consumo, 1990. Cap. 3 y 4.
- Robbins SP. Organizational behavior. Concepts, controversies and applications. Englewood Cliffs: Prentice Hall, 1989. Cap. 17.
- Saturno PJ, Imperatori E, Corbella A (eds.). Evaluación de la calidad asistencial en Atención Primaria. Experiencias en el marco de la cooperación ibérica. Madrid: Ministerio de Sanidad y Consumo, 1990.
- Saturno PJ. Estrategias para la puesta en marcha de actividades de evaluación y mejora de la calidad asistencial en Atención Primaria (I). Consideraciones generales. Centro de Salud 1993; 1(9): 586-592.
- Savater F. Ética como amor propio. Barcelona: Ariel, 2008.
- Schein EH. Organizational culture and leadership. San Francisco: Jossey-Bass, 2004.
- Scholtes PR, et al. The team handbook. Madison: Joiner Associates, 1988.
- The Conceptual Framework for the International Classification for Patient Safety. Final technical report. Geneva: WHO, 2009. Disponible en: http://www.who.int/patientsafety/taxonomy/icps_full_report.pdf (acceso: febrero de 2010).

Capítulo 2

Identificación y priorización de oportunidades de mejora

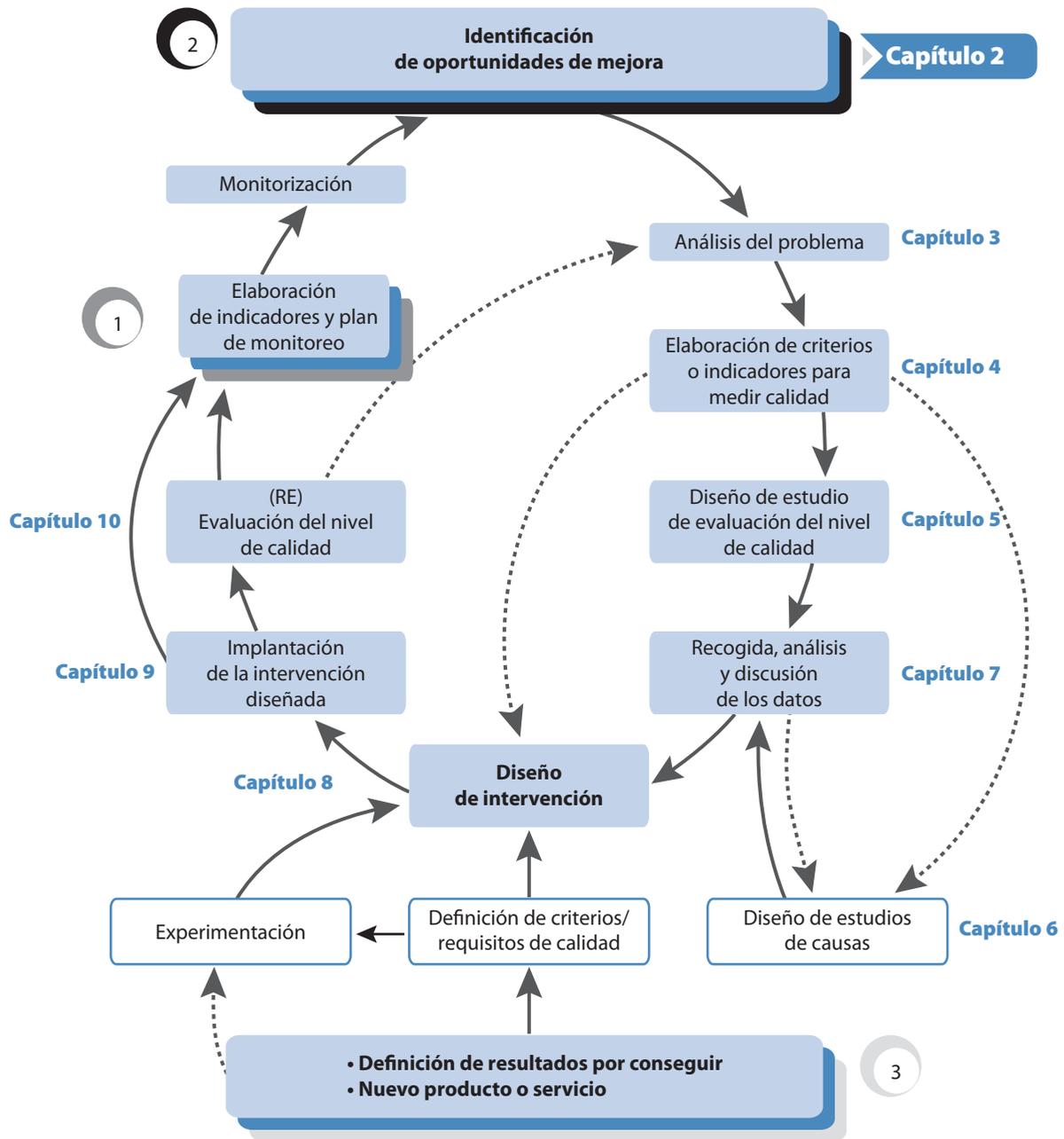
Contenidos generales y objetivos específicos

1. Introducción
2. Concepto de problema de calidad u oportunidad de mejora
3. Identificación de la oportunidad de mejora
4. Normas generales para la selección de una oportunidad de mejora
5. Métodos de identificación de oportunidades de mejora. Consideraciones generales
6. Métodos para la identificación de oportunidades de mejora que no precisan datos
7. Métodos para la identificación de oportunidades de mejora mediante datos
8. Inicio práctico del ciclo de mejora
9. El cartel narrativo: un apoyo para la realización y el seguimiento del ciclo de mejora

Anexo 2.1. Instrucciones para la elaboración de un cartel narrativo de un ciclo de mejora

Anexo 2.2. Instrucciones para la realización de la técnica de grupo nominal

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Puntos de partida



Vía de acción normal



Vía contingente o alternativa



1 Monitorización



2 Ciclo de mejora



3 Diseño

Contenido general

Este capítulo versa sobre el primer paso de los ciclos de mejora: la selección y descripción operativa del problema u oportunidad de mejora que se evaluará.

Se hace una revisión de diversos métodos utilizables, ilustrándolos con ejemplos, y se insta al lector a que seleccione su objetivo de mejora. Para facilitar la realización y seguimiento del ciclo de mejora que se inicia con este capítulo, se aconseja utilizar un cartel narrativo.

Objetivos específicos

1. Describir diferentes métodos aplicables para la identificación de problemas de calidad u oportunidades de mejora
2. Escoger la metodología de identificación de problemas más adecuada a su entorno de trabajo
3. Resumir los resultados de la lluvia de ideas de un grupo de trabajo
4. Dirigir la realización de un Grupo Nominal
5. Distinguir diferentes métodos de priorización de problemas
6. Aplicar matrices de decisión a la priorización de problemas
7. Establecer prioridades basándose en datos
8. Identificar en qué formas puede tenerse en cuenta al usuario para la identificación de oportunidades de mejora
9. Definir qué es un problema de calidad u oportunidad de mejora
10. Describir un problema de calidad en términos operativos concretos
11. Diferenciar problemas genéricos y específicos
12. Diferenciar problemas estructurales de problemas relativos a las actividades y resultados
13. Escoger la oportunidad de mejora que va a ser sometida a evaluación

Identificación y priorización de oportunidades de mejora

*“Un defecto puede ser un tesoro. El descubrimiento de la imperfección nos brinda la posibilidad de reducir el espacio que nos separa de la excelencia”
(citado por Berwick DM)*

1. Introducción

Cualquier actividad de evaluación de la calidad debe comenzar por plantearse ¿Qué es lo que se quiere mejorar? De esta forma iniciamos el ciclo de evaluación: con la identificación de la oportunidad de mejora. Probablemente ésta sea la fase más importante. Una elección acertada será crucial para que tengamos éxito en el proyecto. De hecho, la decisión que tomemos en este momento condicionará el desarrollo metodológico de las siguientes etapas del ciclo (análisis del problema, evaluación e intervención). Si la elección no es acertada, las siguientes fases no se podrán desarrollar con la agilidad requerida y sin tropiezos. Peor aún, puede ocurrir, en un caso extremo, que el desarrollo del ciclo de mejora entre en una vía sin salida. Por tanto es conveniente, sobre todo en las primeras experiencias, que para identificar la oportunidad de mejora reflexionemos sobre las implicaciones de su elección, priorizando las que sean más sencillas en su desarrollo a lo largo del ciclo de mejora y, sobre todo, en las que vislumbremos la factibilidad de su solución.

Es posible que la idea de mejorar en conjunto la calidad de toda la organización pueda parecernos atractiva. Sin embargo, una evaluación global puede presentar numerosas dificultades prácticas, especialmente cuando se aplica a una institución en la que se llevan a cabo numerosos procesos asistenciales y administrativos para resolver condiciones clínicas con requisitos de calidad muy diferenciados, como obviamente ocurre en la mayoría de los centros sanitarios. Dado que la revisión simul-

tánea de todos los servicios prestados es generalmente impracticable como primer paso e incluso es desaconsejable afrontar varias oportunidades de mejora a la vez, es preferible que seleccionemos solamente una. Lo anterior es algo que queremos hacer especialmente bien o que es inadmisibles que se haga mal y que sospechamos que hay una diferencia importante entre la atención proporcionada y la que sería razonable prestar. Las ventajas de esta manera de proceder son numerosas. Focalizamos la evaluación de la calidad en un aspecto concreto y tangible, facilitando el compromiso de los profesionales con la mejora de la calidad. Además, en ocasiones, la delimitación de una oportunidad de mejora será suficiente para que se promuevan los cambios que conducen a la resolución del problema, sobre todo cuando sus causas son evidentes.

En este capítulo revisaremos una serie de herramientas conducentes a delimitar y priorizar oportunidades de mejora. Veremos que tenemos diversas técnicas para su identificación. La decisión de utilizar un método determinado debe condicionarse a las características de cada centro y principalmente a la factibilidad de su empleo. En este sentido, vamos a subrayar la utilización de la lluvia de ideas, la técnica del grupo nominal y la matriz decisional. Esta elección se debe a que estos métodos no requieren un contexto especial de aplicación (no necesitan fuentes de datos adicionales), son fáciles de comprender e involucran a los que quieren evaluar y mejorar la calidad. Son métodos versátiles que pueden adaptarse a la mayoría de las situaciones.

Este capítulo aborda el inicio del ciclo de mejora o ciclo de evaluación: la identificación de la oportunidad de mejora; y se revisan diversos métodos para llevarla a cabo.

2. Concepto de problema de calidad u oportunidad de mejora

Podemos proponer una oportunidad de mejora si percibimos o intuimos que alguna de las características del servicio que ofrecemos no está como creemos que debe de estar. Desde el punto de vista cuantitativo la podemos definir como una diferencia, que potencialmente se puede reducir, entre lo que es razonable que exista (estructura), se haga (proceso) u ocurra (resultado), con lo que existe, se hace o se proporciona en la práctica habitual. Por ejemplo, podemos identificar una oportunidad de mejora cuando percibimos que la captación de hipertensos es muy baja en relación a lo que podríamos esperar, sospechamos que el seguimiento de las dislipemias se realiza de forma esporádica, intuimos que se utilizan antibióticos en la mayoría de los pacientes diagnosticados con infección respiratoria alta, vemos que es demasiado frecuente el que no esté disponible la historia clínica a la hora de la consulta, etc. Una oportunidad de mejora es una divergencia entre la situación actual y la que queremos que exista. ¿Cómo proceder para identificarlas en nuestro entorno? Hay una serie de técnicas que sirven para hacerlas explícitas y que veremos a continuación.

La oportunidad de mejora es cualquier situación en la que se observan diferencias, que se pueden reducir, entre lo óptimo y la realidad.

3. Identificación de oportunidades de mejora

Es evidente que para determinar una posible divergencia entre la atención proporcionada y la que se debería proporcionar, tenemos que haber establecido previamente un marco de referencia, en el que se reflejen de forma explícita y consciente las características que, a nuestro juicio y en función de las necesidades del usuario, debería reunir la atención sanitaria para cumplir con la finalidad para la cual ha sido concebida y con el menor coste posible. Este marco de referencia corresponde al concepto de calidad de los servicios sanitarios que ya habremos establecido previamente, adoptando o adaptando las diversas definiciones posibles.

Sin embargo, este marco de referencia puede resultar excesivamente ambiguo y poco operativo para concretar una oportunidad de mejora. Es preferible que se reflexione acerca de la gran variedad de *productos* o servicios

prestados al usuario y definir a continuación sus respectivos requisitos de calidad. Por ejemplo, la atención de pacientes crónicos con hipercolesterolemia es un producto asistencial con unos requisitos de calidad que le confieren una identidad propia y que le diferencian de otros productos asistenciales, como puede ser el caso de la atención al paciente con cervicalgia, la atención a diabéticos, a la mujer embarazada o la gestión de la cita previa.

Una vez que llegamos a este nivel de especificación debe plantearse para ese determinado proceso o servicio que nos preocupa especialmente ¿Qué tipo de atención es razonable que se proporcione? ¿Qué tipo de atención estamos proporcionando? y ¿Qué diferencia suponemos o percibimos que hay entre ambas? Esta aproximación a los elementos básicos y específicos de la atención sanitaria es probablemente más adecuada para apreciar las posibles divergencias entre lo que es y lo que debería ser y permitirá delimitar con más facilidad diversas oportunidades de mejora de la calidad asistencial que pueden coincidir completamente -perspectiva integral- o en parte -perspectiva parcial- con uno o varios servicios de la actividad asistencial. Si atañe a un solo servicio le vamos a llamar problema específico, mientras que si afecta a más de un servicio le llamaremos problema genérico.

Con estas reflexiones, espontáneamente o con métodos estructurados, es posible identificar un listado normalmente extenso de oportunidades de mejora; sin embargo, para empezar un ciclo de mejora vamos a seleccionar una, y es conveniente que prestemos atención a una serie de normas generales que nos pueden ayudar a que la selección sea adecuada.

4. Normas generales para la selección de una oportunidad de mejora

Cualquier oportunidad de mejora es válida desde el punto de vista operativo siempre que su desarrollo metodológico sea posible (puede analizarse y medirse sin dificultad) y conduzca a su resolución. Para valorar esta validez debemos considerar tanto factibilidad de su resolución como las características intrínsecas del problema, que pueden determinar la mayor o menor garantía de éxito en la mejora de la calidad, y con base en, ello priorizar y seleccionar una.

Las oportunidades de mejora se identifican más fácilmente si pensamos en servicios concretos y sus requisitos de calidad.

4.1 Factibilidad de la resolución del problema de calidad

La identificación y selección de una oportunidad de mejora no debe consistir exclusivamente en revelar o denunciar una atención incorrecta sino tener como objetivo corregirla. Por tanto, es básico que se valoren, al mismo tiempo que se identifica la oportunidad de mejora, las posibilidades de éxito de esa evaluación, en el sentido de que sea factible la mejora. En primer lugar tendríamos que averiguar si existe una relación evidente entre las posibles soluciones y nuestra capacidad de intervención. En la mayoría de las ocasiones, la capacidad de resolver los problemas relacionados con aspectos estructurales de tipo cuantitativo (falta de personal, recursos insuficientes, horario inadecuado, etc.) dependen en gran medida de organismos o de niveles jerárquicos distintos a los que han identificado el problema. Por tanto, en estos casos el margen de maniobra de resolución del problema será escaso y la identificación del problema se limitará con mucha probabilidad a una simple denuncia de la situación. Uno de los errores más comunes en los primeros pasos en la mejora de la calidad consiste en plantear

la oportunidad de mejora en términos de “falta de...”; sin embargo, la ausencia de algo no es evaluable: puede ser una causa del problema o una solución a un problema, pero nunca el problema a evaluar. Es recomendable que descartemos, de entrada, los problemas de estructura en términos de cantidad de recursos humanos o de equipamiento como foco de evaluación.

¿Quiere esto decir que sólo debemos someter a ciclos de mejora a los problemas para los que se intuye una fácil solución? A mediano y largo plazo probablemente no, pero a corto plazo, para las primeras experiencias, rotundamente sí. Es una cuestión de estrategia de implantación de la innovación que supone prestarle una atención explícita y cuantificada a la mejora de la calidad.

4.2 Características intrínsecas de la oportunidad de mejora e implicaciones metodológicas

Desde un punto de vista operativo, además de la visión integral o parcial que ya hemos visto, es útil que clasifiquemos las oportunidades de mejora en función de dos características intrínsecas (figura 2.1): (i) en función de

Cualquier problema de calidad es válido para iniciar un ciclo de mejora siempre que se pueda cuantificar, analizar y resolver.

	Proceso/resultado	Estructura (organizacional)
Específico	Problema tipo I (ejemplo: atención de pacientes con hipercolesterolemia)	Problema tipo II (ejemplo: organización del equipo en la atención de pacientes con hipercolesterolemia)
Genérico	Problema tipo III (ejemplo: control de pacientes con enfermedad crónica)	Problema tipo IV (ejemplo: organización del equipo en la atención de pacientes de la consulta programada)

Figura 2.1. Clasificación de los problemas de calidad

que coincidan con un solo producto o servicio asistencial -problema específico- o se refieran a varios -problema genérico-, tal como hemos apuntado más arriba; y (ii) según afecten a un componente determinado en la prestación de los servicios -problemas de estructura versus problemas de proceso/resultado. La elección de uno de estos tipos de problema debe hacerse de forma consciente, valorando las consecuencias de la elección en relación a su medición y capacidad de resolución. En general, la posibilidad de desarrollar metodológicamente el ciclo de mejora y las posibilidades de éxito de la evaluación son mayores con los problemas de tipo I (específicos y de proceso/resultado) que con los de tipo IV (genéricos y estructurales) pasando por los de tipo II (específicos y estructurales) y tipo III (genéricos y de proceso/resultado).

Problemas específicos mejor que genéricos

Los problemas genéricos presentan, por su tendencia globalizadora, un gran atractivo para resumir diversos servicios asistenciales en una sola evaluación. Sin embargo, precisamente por su tendencia globalizadora, su evaluación puede reflejar espejismos. Cualquier problema se complica y se difumina cuando se trata de forma genérica. Por ejemplo la organización del trabajo en equipo para los atendidos en el centro de salud o en un determinado servicio hospitalario es una oportunidad de mejora de tipo genérico en la que no se puede determinar con precisión sobre qué aspecto hay que incidir para mejorar ¿Sobre la cita a las consultas? ¿El tiempo dedicado a cada consulta? ¿La visita a domicilio? ¿La coordinación con enfermería para la atención a un tipo de paciente concreto? Un problema potencialmente genérico como la organización del trabajo para atender a un tipo de paciente específico es más fácil de comprender y de abordar que la organización del trabajo en general (problema genérico). Del mismo modo, los requisitos del registro de los datos relacionados con el control de un tipo específico de pacientes (problema específico) son más evidentes y fáciles de admitir y concretar que la calidad del registro en la historia clínica en general (problema genérico).

Problemas relacionados con el proceso y/o resultado mejor que los problemas relacionados con la estructura

En la mayoría de los casos, la capacidad de resolución de los problemas estructurales por parte del personal de los centros sanitarios se limita exclusivamente a aquellos que están relacionados con aspectos organizativos. Sin embargo, los aspectos estructurales, aunque son necesarios, no son suficientes para que la calidad aumente. Por ejemplo, la inexistencia de esfigmomanómetro o la falta de fiabilidad del mismo (problema estructural) imposibilitan u ocasionan un déficit en la calidad del diagnóstico y seguimiento de pacientes hipertensos. Sin embargo, si se plantea como un problema de calidad y éste se resuelve, hará posible una atención adecuada, pero no hay ninguna garantía de que esto ocurra realmente. En el mismo sentido, un equipo perfectamente organizado no es sinónimo de que los problemas de salud se atiendan adecuadamente, sin embargo, una mala organización puede ser la causa de defectos en la atención sanitaria. No obstante, un problema de organización que imposibilita que un determinado proceso se pueda llevar a la práctica es equiparable en importancia a los problemas del proceso y/o resultado. Por ejemplo, en un programa de búsqueda de casos de riesgo cardiovascular elevado, un sistema de recepción de pruebas complementarias en el que son frecuentes los extravíos dificultará la adecuada valoración diagnóstica de la población incluida en el programa de prevención; aunque este problema estructural se haría también evidente pero como una de las causas de la mala calidad, si elegimos como oportunidad de mejora la valoración del riesgo cardiovascular en sí mismo. En resumen, si un problema estructural es relevante, es candidato a ser detectado como causa de alteraciones en el proceso o resultado, mientras que la ausencia de problemas estructurales no significa que el proceso y resultado consiguientes no sean defectuosos. A veces los problemas estructurales específicos son buenos candidatos por su fácil abordaje y solución. Pero normalmente no son los mejores. El análisis de problemas de proceso o resultado es más exhaustivo y permite

La factibilidad y facilidad de análisis de los problemas específicos y de proceso/resultados es en general mayor que la de los problemas genéricos y de estructura

determinar diferentes causas, entre las que se encuentran los problemas estructurales subyacentes.

Ya tenemos claro que vamos a seleccionar un problema u oportunidad de mejora que, por cuestión de factibilidad del ciclo de mejora completo, es conveniente que tenga una serie de características. Pero no hemos revisado aún la metodología para identificar oportunidades de mejora para elegir. Esto es lo que vamos a ver a continuación.

5. Métodos de identificación de oportunidades de mejora. Consideraciones generales

Hay muchos métodos y, en principio, cualquier método puede ser válido; al momento de utilizar uno u otro es conveniente ser flexibles y adaptarse a las posibilidades del contexto. Tampoco hay por qué utilizar sólo uno, no son excluyentes. Uno de los principales factores determinantes de nuestra elección del método o métodos será el que dispongamos de datos sobre los servicios que prestamos o no.

Si no disponemos de datos, la fuente de información que nos permite apreciar si hay problemas de calidad será exclusivamente la propia experiencia; el conocimiento que tengamos de la actividad de nuestros centros. Como se puede suponer, la principal limitación de estos métodos que no utilizan datos, entre los que destacan la lluvia de ideas y la técnica del grupo nominal, radica en que su efectividad es directamente proporcional al grado en el que conocemos de forma directa los servicios que se quieren mejorar. Lo que implica la necesidad de que seamos proveedores de esos servicios o que tengamos un contacto directo con los mismos, para que estos métodos se puedan utilizar con provecho. Por otra parte, cuando los utilizamos podemos sospechar la existencia de un problema de calidad pero, aunque es muy probable que éste exista, sólo será una oportunidad de mejora o un problema de calidad potencial, hipotético, del que no tendremos constancia de su magnitud ni de si constituye en realidad un problema hasta que se cuantifique. Sin embargo, las ventajas de los métodos que no precisan datos sino el conocimiento del contexto en que se dan, superan a

sus posibles limitaciones. Si los aplicamos adecuadamente, la oportunidad de mejora identificada se corresponderá con las preferencias de los profesionales evaluados, lo que facilitará el compromiso de éstos con la resolución del problema.

Los métodos que requieren datos, entre los que se encuentran el micromuestreo, la monitorización de indicadores, el análisis de perfiles, y la participación del usuario, no siempre estarán inicialmente al alcance de cualquier centro ya que precisan sistemas de información específicos. Obviamente, serán más precisos en la identificación de oportunidades de mejora que los puramente cualitativos, ya que cuantifican la dimensión del problema.

Vamos a ver con más detalle ambos tipos de métodos, aunque enfatizando los métodos sin datos porque son más versátiles, sencillos y están al alcance de todo el mundo. El cuadro 2.1 contiene una relación de los métodos que vamos a revisar, agrupados según precisen o no datos.

Cuadro 2.1	
Métodos de identificación y priorización de problemas	
1.	Métodos que no requieren datos
	1.1. Lluvia de ideas (<i>brainstorming</i>)
	1.2. Técnica de grupo nominal
	1.3. Matriz decisional
	1.4. Comparación por pares
	1.5. Comunicación de incidencias
2.	Métodos que requieren datos
	2.1. Micromuestreo
	2.2. Monitorización de indicadores
	2.3. Análisis de perfiles estadísticos
	2.4. La participación del usuario

6. Métodos para la identificación de oportunidades de mejora que no precisan datos

6.1. La lluvia de ideas (*brainstorming*)

Vamos a emplear esta técnica principalmente cuando no tenemos mucha experiencia en la toma de decisiones en grupo y cuando el objetivo es sacar a la luz una gran cantidad de

oportunidades de mejora. Sin embargo, hay que considerar que el grado de conclusión de los resultados será menor en relación a la técnica del grupo nominal y que habrá que completar el proceso de identificación de problemas con alguna técnica de priorización como, por ejemplo, la matriz decisional o la técnica del grupo nominal. Además, los resultados pueden polarizarse, reflejando sobre todo los problemas aportados por aquellos miembros del grupo con mayor tendencia a la participación y a la exteriorización de sus impresiones.

El principio básico para llevar a cabo esta técnica consiste en evitar cualquier tipo de valoración o crítica cuando los participantes exponen sus ideas. De esta manera, no se coaccionará la libertad de los participantes ni se dificultará un proceso creativo que puede ser incompatible con un juicio crítico anticipado. Podemos identificar cuatro normas de interés en la ejecución de este método (cuadro 2.2). Tras la introducción, podemos empezar con la propuesta de una oportunidad de mejora, que en un principio pueda parecer banal, con el fin de crear un ambiente distendido que promueva a la participación y favorezca el planteamiento del mayor número de ideas. A partir de este momento, el encadenamiento y la asociación de ideas son la base que facilita una mayor creatividad. Cada participante planteará un solo problema en cada intervención cuando lo considere oportuno y sin que se establezca un límite de intervenciones. Durante los episodios de silencio, el moderador puede aportar nuevas ideas o variaciones sobre el tema. En ocasiones tendremos que completar este método con otras técnicas de consenso, es-

pecialmente cuando se necesiten clarificar los problemas planteados y redefinir aquellos que presenten puntos en común. Es difícil seleccionar una oportunidad de mejora solamente con la lluvia de ideas.

6.2. Técnica de grupo nominal

Podemos aplicarla cuando el grupo no es muy heterogéneo y tiene un tamaño inferior a 10 personas. Con esta técnica normalmente conseguiremos una gran validez en las decisiones realizadas a nivel grupal, a la vez que se favorece la resolución constructiva de los conflictos despersonalizando la confrontación entre los desacuerdos.

Este método se basa en utilizar la reflexión individual para generar ideas y para su priorización. A lo largo de su desarrollo se procura la participación paritaria de todos los componentes del grupo para asegurar que se expresen también las posturas e ideas minoritarias, las cuales quedarían ocultas o no se llegarían a expresar en una discusión abierta. Su desarrollo se realiza de una forma muy estructurada y requiere que se sigan las siguientes etapas:

1. Generación de ideas de forma individual.

Después de plantear en forma de pregunta escrita, concreta y sin ambigüedades la cuestión sobre la que hay que reflexionar (en nuestro caso, las oportunidades de mejora existentes) los participantes escriben sus propuestas, de forma individual, sin que intercambien información y sin que se realicen aclaraciones sobre el tema. Para mayor operatividad es conveniente limitar el número de propuestas por cada individuo, por ejemplo a tres o cinco.

2. Registro del listado individual de ideas.

Para ello solicitaremos a cada participante que de forma rotatoria exponga su listado de problemas (oportunidades de mejora) por turno de intervención para registrarlos en un listado común.

3. Clarificación de ideas.

Los problemas planteados en el listado común se someten a una discusión que tendrá como principal objetivo aclarar su significado, evitando duplicaciones y malentendidos. Posteriormente se puede volver a elaborar el listado eliminando los

La lluvia de ideas favorece la creatividad y la identificación de una gran cantidad de oportunidades de mejora, pero es difícil que sirva por sí sola para seleccionar una.

La técnica del grupo nominal favorece la participación paritaria de todo el grupo en la identificación de la oportunidad de mejora.

Cuadro 2.2

Normas para el desarrollo de la lluvia de ideas

1. Exposición del objetivo de la reunión por parte del moderador.
2. Repaso de las reglas del método: no criticar las ideas propuestas.
3. Planteamiento, por parte del animador, de un ejemplo que puede parecer banal, pero que favorece un ambiente distendido y la generación de ideas.
4. Exposición, por cada participante, de una sola idea a la vez cuando considere oportuno y sin limitación en su número.

conceptos que se encuentren repetidos y concretando o matizando aquellos que sean ambiguos o que no respondan completamente a la cuestión planteada.

4. Priorización individual de las ideas (votación preliminar). Durante esta fase se valoran de forma individual e independiente (sin discusión de grupo) las oportunidades de mejora del listado definitivo, otorgándoles una puntuación según se consideren más o menos relevantes. De esta manera se evitarán las presiones de los compañeros, y se garantizará que la toma de decisiones refleje lo que opinan todos los participantes. Obtendremos la decisión del grupo calculando el promedio de las priorizaciones individuales. La puntuación es el método de priorización más simple y el que se emplea habitualmente en el grupo nominal (por ejemplo de 0 a 5 puntos para cada problema, repartiendo un número de votos fijo entre todos ellos, o de cualquier otra forma que acuerde el grupo). Sin embargo, en este paso pueden utilizarse también otros métodos complementarios como la matriz decisional que veremos más adelante.

5. Discusión del voto preliminar. Cada participante expresa su votación y se suman todas ellas para obtener la priorización del grupo. La decisión del grupo puede someterse a una nueva valoración en la que se intenta discutir las votaciones obtenidas. Se consideran especialmente los apartados con patrones de votación heterogéneos, o puntuaciones extremas (muy elevadas o muy bajas). En estos casos es importante que se descarte la existencia de factores artificiales, como por ejemplo la información insuficiente o parcial, así como interpretaciones incorrectas sobre la cuestión planteada. Si se ve oportuno, tras esta discusión puede procederse a la votación final.

6. Voto final. Con esta etapa se finaliza el proceso de toma de decisiones procediéndose de forma análoga a la 4ª etapa. El problema u oportunidad de mejora más votado será el elegido por el grupo para su evaluación y mejora.

El anexo 2.2 contiene instrucciones que pueden servir de guía para la realización de la técnica del grupo nominal

6.3. Otros métodos de selección de oportunidades de mejora: notificación de incidentes, comparación por pares y matriz decisional o parrilla de análisis de criterios múltiples

Hay otros métodos, enfocados sobre todo a la priorización, que establecen unos criterios de valoración para evaluar la relevancia del problema. Para definir estos criterios las indicaciones de la *Joint Commission on Accreditation of Health Care Organizations* (JCAHO) pueden ser orientativas en el sentido de seleccionar procesos que sean conjunta o separadamente muy frecuentes, de alto riesgo para los usuarios o con tendencia reconocida a ser problemáticos. Además se pueden añadir la dependencia interna de la solución y el coste de la misma. Cada centro determinará la adopción, adaptación o incorporación de estos u otros criterios de priorización, así como su grado de relevancia.

Los criterios que decidamos para valorar la oportunidad de mejora los podemos aplicar de forma implícita o bien explícita, como por ejemplo en la comparación por pares y en la matriz decisional respectivamente. Esta última normaliza las valoraciones realizadas por los diversos participantes en el proceso de priorización, a diferencia de lo que ocurre con la comparación por pares que al ser implícita es más subjetiva. Ambos métodos los revisaremos más a detalle a continuación.

Otro método sencillo y de utilidad, sobre todo cuando ya tenemos una estructura organizativa para las actividades de gestión de la calidad, es el establecimiento de un sistema de recogida de incidencias o problemas (problem report) identificados por los profesionales del centro. Estos problemas son considerados como posibles oportunidades de mejora.

6.4. La comparación por pares

Con este método el proceso de valoración se realiza concentrándose en dos problemas como máximo de forma simultánea, comparando un problema con el otro y eligiendo aquel que se considere más importante. No es imprescindible, por tanto, que se apliquen criterios explícitos de valoración. Podemos distinguir en su desarrollo las siguientes etapas (el esquema final puede verse en la figura 2.2): (i) elaboración de

La técnica del grupo nominal es muy estructurada y se realiza siguiendo una serie de pasos determinados.

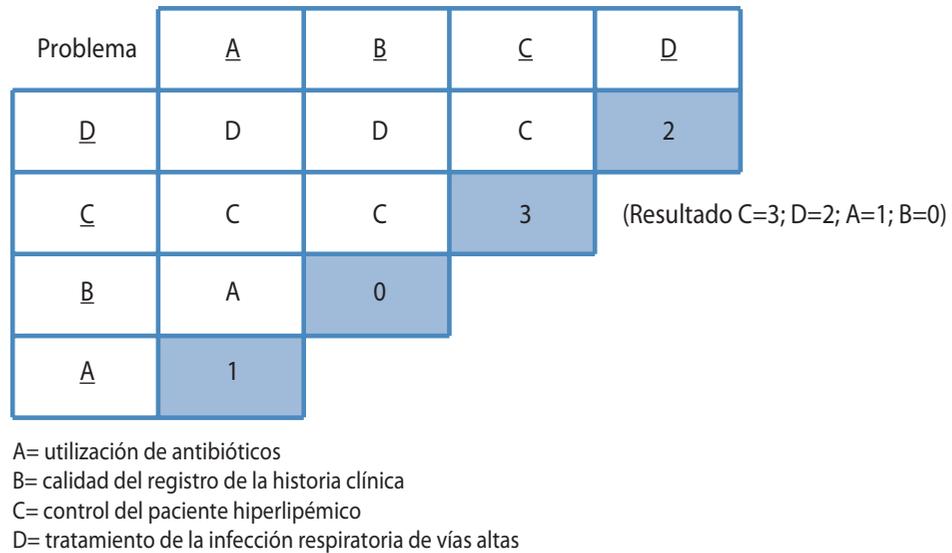


Figura 2.2. Priorización de problemas de calidad mediante el método de comparación por pares

una tabla en la que se detallan como cabeza de las filas y las columnas todos los problemas a considerar; (ii) confrontación de todas las parejas posibles, combinando los problemas de dos en dos, al cruzar filas y columnas; (iii) selección, dentro de cada pareja, del problema más importante. Ese problema se anota en la celda correspondiente al cruce donde se encuentra la pareja comparada. Cada pareja se compara sólo una vez; (iv) cómputo final, contabilizando el número de veces en los que ha sido elegido cada problema durante la fase anterior; y (v) priorización de los problemas clasificándolos en orden decreciente a la puntuación obtenida. La figura 2.2 esquematiza este proceder aplicándolo a la priorización de cuatro problemas.

6.5. Matriz decisional de criterios múltiples o parrilla de análisis

Esta herramienta tiene dos modalidades de aplicación en función de que los criterios empleados se consideren igual de importantes o bien se jerarquicen.

En el primer caso la parrilla de análisis adopta una forma matricial en la que se incluyen en la primera columna los distintos pro-

blemas identificados y en la primera fila los criterios de valoración (figura 2.3). Los problemas de calidad se valoran de forma separada concediéndoles una puntuación en cada uno de los criterios para contabilizar posteriormente el total de puntos obtenidos. El nivel de priorización se establece ordenando los problemas de calidad según la puntuación obtenida.

En el segundo caso, previa jerarquización o priorización de los criterios decisionales, se diseña un gráfico de forma arborescente en el que los criterios de valoración se representan como nódulos sucesivos y su valoración positiva o negativa se representa como un segmento superior o inferior respectivamente, que se inicia en ese nódulo y conecta con el siguiente (figura 2.4). El criterio más importante se sitúa en el primer nódulo, continuándose con los demás criterios en función del orden decreciente de importancia. Cuando se valora un problema se sigue el recorrido determinado por la valoración positiva o negativa de los criterios y una vez que se llega al último se señala su nivel de prioridad.

La matriz decisional puede utilizarse como método de priorización individual en el paso 4 de la técnica de grupo nominal.

La comparación por pares prioriza con criterios implícitos, subjetivos.

Criterios de decisión (máximo=5; mínimo=1)					
Problema	¿Afecta a muchos pacientes?	¿Supone un riesgo grave para la salud?	¿Dependencia interna de la posible solución?	¿Es una solución barata?	Total
A	++	++	+++	++++	11
B	++	+	+++	++++	10
C	++	+++++	+++	++++	14
D	+++++	+	+++	++++	12

A= uso no racional de antibióticos
 B= calidad del registro de la historia clínica
 C= control del paciente hiperlipémico
 D= tratamiento de la infección respiratoria de vías altas

Figura 2.3. Parrilla de decisión mediante criterios univalentes

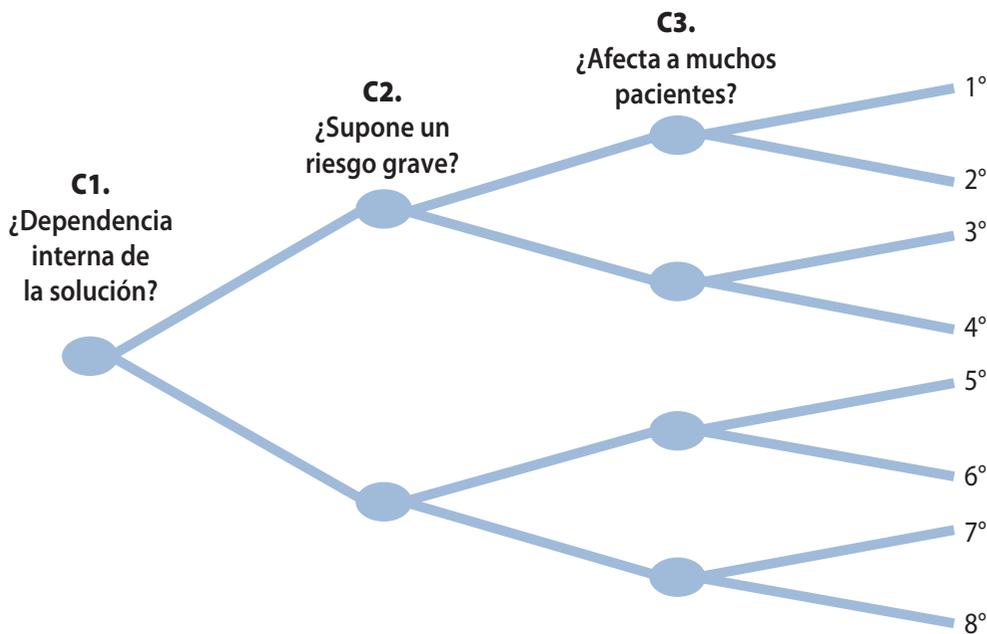


Figura 2.4. Parrilla de análisis aplicando criterios jerarquizados en orden de importancia (C1>C2>C3) (la línea continua corresponde a la valoración positiva del criterio)

La combinación de lluvia de ideas, grupo nominal y matriz decisional sin jerarquización de criterios (figura 2.3) es probablemente una de las más frecuentes y mejores bases para reali-

zar la identificación y priorización de oportunidades de mejora, contando exclusivamente con el conocimiento que se tiene del entorno de trabajo en el que se prestan los servicios a evaluar.

7. Métodos para la identificación de oportunidades de mejora mediante datos

El uso de este tipo de métodos está condicionado por la disponibilidad de fuentes de información que suministren los datos necesarios para aplicarlos. La posibilidad de identificar oportunidades de mejora puede ser más amplia con el micromuestreo y los métodos de participación del usuario mientras que la monitorización y el análisis de perfiles se suelen circunscribir a aspectos asistenciales concretos y previamente seleccionados.

7.1. Micromuestreo de registros

Este método requiere que se registren los datos de la actividad asistencial. Cuando disponemos de estos registros (historias clínicas, documentos de derivación, prescripciones) podemos hacer una revisión rápida de una pequeña parte de los mismos (por ejemplo 15 o 20) para detectar áreas potencialmente problemáticas. Para aumentar la validez de este método se suele revisar cada registro por dos profesionales con experiencia en la materia. Estos valoran si a su juicio existe alguna actuación incorrecta en la atención proporcionada al paciente, comparando y resumiendo posteriormente sus anotaciones, a partir de las cuales se puede elaborar un listado de problemas. Pueden buscarse problemas genéricos (de antecedentes, prevención, utilización de analítica, etc.) o específicos, si hacemos el muestreo sobre un tipo concreto de pacientes.

7.2. Monitorización de indicadores

Los sistemas de monitorización de la calidad tienen como objetivo principal identificar problemas de calidad. Por tanto, si tenemos un sistema de medición periódica de indicadores sobre aspectos asistenciales concretos (tiempos de espera, control del paciente diabético, del paciente con cervicalgia, infarto agudo de miocardio, etc.) podemos identificar problemas relacionados con ese área asistencial. Cuanto más amplio sea el listado de indicadores utilizado, mayor será el espectro de problemas y servicios sobre los que será posible identificar la existencia de problemas, pero siempre va a

ser más restrictivo que los métodos que no precisan datos.

7.3. Análisis de perfiles estadísticos

Cuando la medición de un indicador se realiza entre distintos proveedores (facultativos, centros sanitarios, etc.), los resultados se pueden representar mediante un gráfico de barras. Cada barra refleja el nivel alcanzado por cada uno de los proveedores en la medición del indicador (figura 2.5). Una amplia variabilidad en la distribución de los perfiles sugiere la existencia de un problema de calidad en alguna o varias áreas asistenciales que se encuentran relacionadas con el indicador valorado, especialmente en aquellos proveedores que se encuentran alejados en más de una o dos desviaciones estándar en relación a la media. Este método se ha empleado con frecuencia para identificar posibles oportunidades de mejora en la prescripción de medicamentos, estancia media para la misma patología y utilización de recursos en general. En el campo de la prescripción, se han utilizado indicadores, entre otros recursos, como la proporción de medicamentos de utilidad terapéutica baja, la proporción de antibióticos recetados en atención primaria pertenecientes al grupo de las quinolonas o las cefalosporinas, etc. Obviamente, precisa de una priorización previa de los aspectos a medir. La figura 2.5 es un ejemplo de análisis de perfiles. Una variante es la representación de las medias y los percentiles, o la estimación con sus intervalos de confianza.

7.4. La participación del usuario en la identificación de problemas

El paciente, usuario del servicio y población en general son el destinatario final de la atención sanitaria. Su papel en la identificación de problemas de calidad es esencial en aquellos aspectos que están relacionados con el incumplimiento de sus expectativas, a la vez que constituye una fuente de información adicional para identificar oportunidades de mejora de procesos asistenciales que no se registran habitualmente. Su implicación en la identificación de oportunidades de mejora por parte de los centros asistenciales se puede articular me-

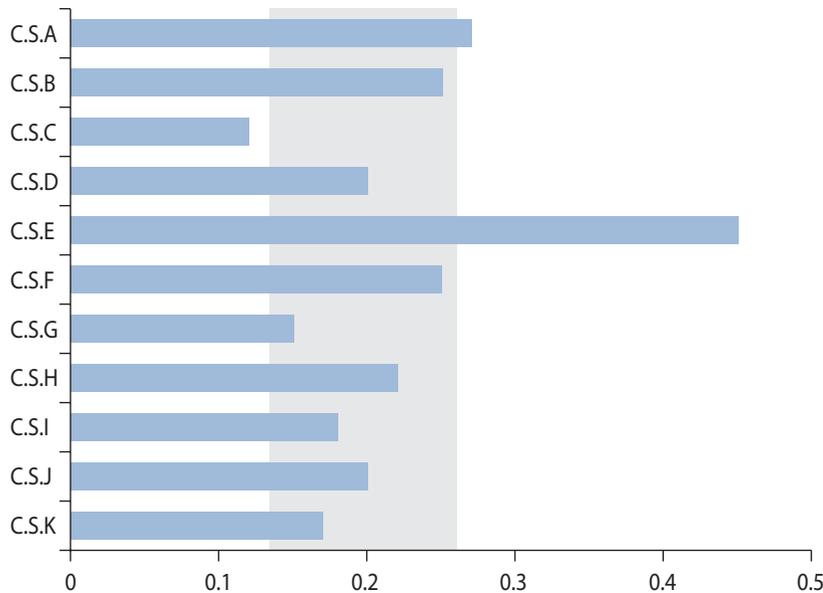


Figura 2.5. Análisis de perfiles estadísticos

Ejemplo: Proporción de quinolonas y cefalosporinas en relación al total de antibióticos recetados en 11 centros de salud. El área sombreada incluye a la media \pm una desviación estándar

diante el análisis de las quejas y reclamaciones, las encuestas de satisfacción, calidad percibida y los informes del usuario. Todas ellas precisan de una infraestructura previa para su realización y aprovechamiento. Resumimos aquí lo fundamental de estas herramientas como métodos de identificación de problemas.

7.5. Análisis de quejas y reclamaciones

Las ventanillas o los libros de recepción de quejas y reclamaciones son un medio clásico de identificación de problemas, aunque con frecuencia se haya desvirtuado su naturaleza, considerándolos principalmente como un sistema que es obligatorio tener y eventualmente punitivo. El número de reclamaciones recibidas en un centro sanitario puede ser un reflejo de su calidad y de las facilidades prestadas al usuario para participar en la identificación de problemas. En ocasiones tendremos que valorarlas con precaución ya que pueden responder a una interpretación errónea y no señalar en realidad un problema de calidad, pero en general nos van a ser de utilidad para identi-

ficar problemas relacionados con los aspectos que más valora el usuario, y pueden ser la punta del iceberg de problemas que debemos solucionar.

7.6. Encuestas de satisfacción y calidad percibida

Con las encuestas de satisfacción y calidad percibida podemos identificar problemas similares a los señalados en el análisis de quejas y reclamaciones (no cumplimiento de las expectativas del usuario). Pero las ventajas sobre este último radican en que: (i) si la encuesta es metodológicamente correcta, identifica una mayor cantidad de problemas (relacionándolos directamente con la satisfacción) y (ii) refleja los puntos de vista de una muestra representativa de pacientes, de manera que se evita el sesgo de incluir sólo la opinión de los pacientes que expresan su insatisfacción a iniciativa propia. En definitiva, las encuestas de satisfacción y calidad percibida son en general una fuente de información más adecuada para sacar a la luz problemas relacionados con el trato recibido, el

El micromuestreo y la participación del usuario pueden identificar una gran variedad de oportunidades de mejora, pero la monitorización y el análisis de perfiles se focalizan en aspectos asistenciales.

contexto en el que se desarrolla la asistencia y también problemas de tipo organizacional.

7.7. El informe del usuario

De forma complementaria a las encuestas de satisfacción y calidad percibida, e incluso de los registros clínicos en algunos de sus aspectos, se puede preguntar al propio usuario sobre aspectos de la asistencia recibida que él conoce (¿Le informaron cuántas veces al día tenía que tomar esa medicación?, ¿Cuándo le tomaron la tensión arterial por última vez? Durante su estancia en el hospital, ¿le prestaron ayuda para su aseo siempre que la necesitó?). La principal ventaja de este método consiste en revelar áreas-problema de aspectos asistenciales que habitualmente no se registran, pudiendo ser, en ocasiones, un método incluso más eficiente de recopilación de datos en la identificación de problemas que la revisión de registros.

8. Inicio práctico del ciclo de mejora

Hemos revisado una serie de métodos para poder identificar y priorizar oportunidades de mejora, tanto si tenemos como si no tenemos datos; y hemos señalado los métodos sin datos como los más sencillos y de aplicación universal. Con este bagaje estaríamos en condiciones de elegir la oportunidad de mejora utilizando el método más conveniente. En principio podríamos sondear en el conjunto global de la atención prestada y evaluar aquel servicio que pareciera prioritario. Sin embargo, es conveniente conocer en la práctica cómo pueden funcionar algunas de las herramientas que hemos visto en este capítulo. Vamos a seleccionar dos: la técnica de grupo nominal y la matriz decisional de criterios múltiples. Animamos por tanto a realizar un ejercicio práctico sobre identificación de problemas para mejorar. Si se va a realizar en grupo, se aconseja utilizar la técnica de grupo nominal, usando puntuación simple para la priorización o bien la matriz decisional; si el trabajo es individual, puede utilizarse la matriz decisional directamente.

9. El cartel narrativo: un apoyo para la realización y el seguimiento del ciclo de mejora

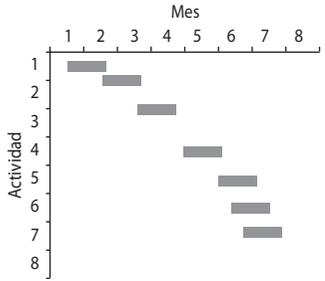
El *story board* o cartel narrativo del proyecto de mejora es una representación paso a paso del ciclo de mejora. Aunque puede empezar a utilizarse a cualquier altura dentro del ciclo de mejora, lo más lógico es establecerlo una vez decidido el problema a evaluar y mejorar o, como mínimo, una vez decidida la intervención para mejorar. Los elementos del cartel completo se pueden ver en la figura 2.6. Al principio estarán los espacios en blanco y se irán ilustrando y llenando a medida que vaya avanzando el proyecto. Para cada paso del ciclo de mejora se va informando de las actividades clave y sus herramientas utilizadas. El que esté todo ello expuesto en un lugar visible fomenta la responsabilidad y compromiso del grupo de mejora encargado del proyecto, a la vez que hace posible y fácil recibir comentarios y sugerencias de otros profesionales del centro.

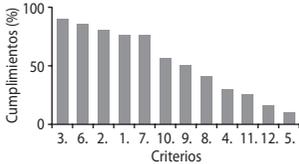
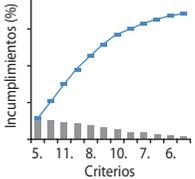
Esta herramienta de ayuda a la implementación de los proyectos se originó de forma casi casual en la industria (concretamente la popularizó *Florida Power and Light*, la primera compañía americana que ganó el Premio Deming), pero se ha extendido a todos los sectores, incluido el sanitario, por su sencillez y utilidad, fomentando la transparencia de las actividades de mejora de la calidad y, como ya hemos apuntado, el compromiso de los grupos de mejora implicados. Adicionalmente, es una herramienta total y absolutamente compatible con todas las demás; de hecho, cualquier método que se utilice en cualquiera de los pasos del ciclo de mejora puede y debe verse reflejado en el cartel, que tiene así, además, una función educativa para todo el personal del centro.

Se incluyen como anexo 2.1 las instrucciones para su elaboración, adaptadas para uso en países latinoamericanos por el proyecto de Garantía de Calidad (*QA Project*) que se realizó con fondos de la AID (*Agency for International Development*) norteamericana, y gestionado por la URC (*University Research Corporation*).

Los carteles narrativos fomentan la transparencia de las actividades de mejora de la calidad, el compromiso de los grupos de mejora

Narración de EMC

Definición del problema	Equipo	Foto	Cronograma
	Nombres: Roles:		

Análisis del problema	<p>Lista de criterios</p> <ol style="list-style-type: none"> 1. Diagnóstico correcto 2. Antecedentes familiares 3. Antecedentes personales 4. Consumo de tabaco 5. Consumo de alcohol 6. Peso y talla y/o IMC 7. Auscultación cardíaca 8. Exploración abdominal 9. Pulsos periféricos 10. Soplos carotídeos 11. Edemas 12. Fondo de ojo <p style="text-align: center;">Criterios</p>	<p>Listado de chequeo</p> <table border="1"> <thead> <tr> <th>Questionario</th> <th>si/no</th> </tr> </thead> <tbody> <tr><td>1. xxxxxxxx? A B C</td><td>xxxxxxxx</td></tr> <tr><td>2. xxxxxxxx? A B C</td><td>xxxxxxxx</td></tr> <tr><td>3. xxxxxxxx? A B C</td><td>xxxxxxxx</td></tr> <tr><td>4. xxxxxxxx? A B C</td><td>xxxxxxxx</td></tr> <tr><td>5. xxxxxxxx? A B C</td><td>xxxxxxxx</td></tr> </tbody> </table> <p style="text-align: center;">Instrumentos</p>	Questionario	si/no	1. xxxxxxxx? A B C	xxxxxxxx	2. xxxxxxxx? A B C	xxxxxxxx	3. xxxxxxxx? A B C	xxxxxxxx	4. xxxxxxxx? A B C	xxxxxxxx	5. xxxxxxxx? A B C	xxxxxxxx	<table border="1"> <thead> <tr> <th>Criterio</th> <th>Núm. incumplimientos (frecuencia absoluta)</th> <th>% frecuencia (relativa)</th> <th>Frecuencia acumulada</th> </tr> </thead> <tbody> <tr><td>5. Consumo de alcohol</td><td>54</td><td>15.8</td><td>15.8</td></tr> <tr><td>12. Fondo de ojo</td><td>51</td><td>14.9</td><td>30.7</td></tr> <tr><td>11. Edemas</td><td>45</td><td>13.2</td><td>43.9</td></tr> <tr><td>4. Consumo de tabaco</td><td>42</td><td>12.3</td><td>56.1</td></tr> <tr><td>8. Exploración abdominal</td><td>36</td><td>10.5</td><td>66.7</td></tr> <tr><td>9. Pulsos periféricos</td><td>30</td><td>8.8</td><td>75.4</td></tr> <tr><td>10. Soplos carotídeos</td><td>27</td><td>7.9</td><td>83.3</td></tr> <tr><td>1. Diagnóstico correcto</td><td>15</td><td>4.4</td><td>87.7</td></tr> <tr><td>7. Auscultación cardíaca</td><td>15</td><td>4.4</td><td>92.1</td></tr> <tr><td>2. Antecedentes familiares</td><td>12</td><td>3.5</td><td>95.6</td></tr> <tr><td>6. Peso y talla y/o IMC</td><td>9</td><td>2.6</td><td>98.2</td></tr> <tr><td>3. Antecedentes personales</td><td>6</td><td>1.8</td><td>100.0</td></tr> <tr><td>Total</td><td>342</td><td>100</td><td></td></tr> </tbody> </table> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;">  <p>Cumplimientos (%)</p> </div> <div style="text-align: center;">  <p>Incumplimientos (%)</p> </div> </div> <p style="text-align: center;">Presentación y análisis de datos</p>	Criterio	Núm. incumplimientos (frecuencia absoluta)	% frecuencia (relativa)	Frecuencia acumulada	5. Consumo de alcohol	54	15.8	15.8	12. Fondo de ojo	51	14.9	30.7	11. Edemas	45	13.2	43.9	4. Consumo de tabaco	42	12.3	56.1	8. Exploración abdominal	36	10.5	66.7	9. Pulsos periféricos	30	8.8	75.4	10. Soplos carotídeos	27	7.9	83.3	1. Diagnóstico correcto	15	4.4	87.7	7. Auscultación cardíaca	15	4.4	92.1	2. Antecedentes familiares	12	3.5	95.6	6. Peso y talla y/o IMC	9	2.6	98.2	3. Antecedentes personales	6	1.8	100.0	Total	342	100	
Questionario	si/no																																																																						
1. xxxxxxxx? A B C	xxxxxxxx																																																																						
2. xxxxxxxx? A B C	xxxxxxxx																																																																						
3. xxxxxxxx? A B C	xxxxxxxx																																																																						
4. xxxxxxxx? A B C	xxxxxxxx																																																																						
5. xxxxxxxx? A B C	xxxxxxxx																																																																						
Criterio	Núm. incumplimientos (frecuencia absoluta)	% frecuencia (relativa)	Frecuencia acumulada																																																																				
5. Consumo de alcohol	54	15.8	15.8																																																																				
12. Fondo de ojo	51	14.9	30.7																																																																				
11. Edemas	45	13.2	43.9																																																																				
4. Consumo de tabaco	42	12.3	56.1																																																																				
8. Exploración abdominal	36	10.5	66.7																																																																				
9. Pulsos periféricos	30	8.8	75.4																																																																				
10. Soplos carotídeos	27	7.9	83.3																																																																				
1. Diagnóstico correcto	15	4.4	87.7																																																																				
7. Auscultación cardíaca	15	4.4	92.1																																																																				
2. Antecedentes familiares	12	3.5	95.6																																																																				
6. Peso y talla y/o IMC	9	2.6	98.2																																																																				
3. Antecedentes personales	6	1.8	100.0																																																																				
Total	342	100																																																																					

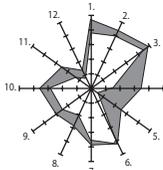
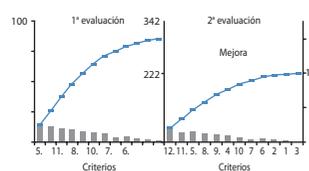
Causas principales	Solución	Resultados												
1. _____ 2. _____	_____ _____ _____	<div style="display: flex; justify-content: space-around;">  <table border="1"> <thead> <tr> <th>Antes</th> <th>Después</th> <th>Manteniendo lo alcanzado</th> </tr> </thead> <tbody> <tr> <td>% _____</td> <td>% _____</td> <td>% _____</td> </tr> <tr> <td>% _____</td> <td>% _____</td> <td>% _____</td> </tr> <tr> <td>Fecha _____</td> <td>Fecha _____</td> <td>Fecha _____</td> </tr> </tbody> </table> </div> <div style="text-align: center; margin-top: 10px;">  <p>1ª evaluación: 342 2ª evaluación: 222 Mejora</p> </div>	Antes	Después	Manteniendo lo alcanzado	% _____	% _____	% _____	% _____	% _____	% _____	Fecha _____	Fecha _____	Fecha _____
Antes	Después	Manteniendo lo alcanzado												
% _____	% _____	% _____												
% _____	% _____	% _____												
Fecha _____	Fecha _____	Fecha _____												

Figura 2.6. Cartel narrativo de un ciclo de mejora C

BIBLIOGRAFÍA

- Delbecq AL, Van de Ven AH, Gustafson DH. Técnicas grupales para la planeación. México DF: Trillas, 1984.
- Palmer RH. Evaluación de la asistencia ambulatoria: principios y práctica. Madrid: Ministerio de Sanidad y Consumo, 1990. Cap 7.
- Pineault R, Daveluy C. La planificación sanitaria: Conceptos, métodos, estrategias. Barcelona, Masson, 1987. Cap. 3.
- Sannazaro PJ, Mills DH. A critique of the use of generic screening in quality assessment. JAMA 1991; 265:1977-1981.
- Saturno PJ. Los métodos de participación del usuario en la evaluación y mejora de la calidad de los servicios sanitarios. Rev Esp Salud Publica 1995; 69:163-175.
- Saturno PJ. Monitorización del uso de medicamentos para mejorar la calidad de la prescripción: problemas, métodos e indicadores. Aten Primaria 1996; 18(6):331-338.

Anexo 2.1

Instrucciones para la elaboración de un cartel narrativo de un ciclo de mejora

■ Cartel narrativo de evaluación y mejora de la calidad ⁽¹⁾

Enunciado del problema

- Describir la oportunidad que existe de mejorar en términos de sus consecuencias. El enunciado del problema debe definir con claridad la diferencia entre la situación deseada y la existente.
- No debe hacer ninguna referencia a la causa del problema; no debe culpar a otros, ni debe expresarse como una solución. Deberá ser evidente dónde comienza y termina el problema, de manera que se marque la diferencia respecto de otros problemas del sistema en general.
- Si se desea, también puede describirse el método o los métodos utilizados para seleccionar un problema o campo para mejorar.

El equipo de EMC (evaluación y mejora de la calidad)

- Identificar a todos los integrantes del equipo, sus funciones en las organizaciones y las funciones que cumplirán en el esfuerzo de mejora de la calidad. Si es posible se puede incluir una fotografía del equipo. También se puede incluir un cronograma general que enumere todas las actividades principales y cuándo se espera iniciarlas y terminarlas. No es necesario seguir exactamente lo que indique el cronograma general, dado que sólo proporciona una guía general para los miembros del equipo.

Análisis del problema

- El análisis del problema debe documentar los esfuerzos realizados por el equipo para determinar cuáles son las causas de fondo del problema. No es necesario incluir todos los elementos que se enumeran a continuación, pero debe mencionarse la causa o las causas de fondo del problema, las que constituirán el punto central de las actividades de mejora.

Diagrama de flujo

- Sirve para mostrar el proceso que existe actualmente e indicar cuáles son los límites del esfuerzo para mejorar la calidad. Indicar los impedimentos o los puntos problemáticos con una flecha. Trazar un círculo alrededor de los campos en los que será necesario recopilar datos.

Diagrama de causa y efecto

- Mostrar el análisis de las causas y efectos utilizando un diagrama de causa y efecto. Marcar aquellas causas sobre las que deben recogerse datos.
- Recordar que se trata de causas hipotéticas hasta que estén fundamentadas con datos.

⁽¹⁾ Batalden P, Gillen T: "Hospitalwide Quality Improvement Storytelling", Quality Resource Group, Hospital Corporation of America, Nashville, Noviembre de 1989.

.../continuación.

Indicadores o criterios principales

- Criterios o indicadores que cuantifican la diferencia entre el desempeño real y el deseado, así como los que determinen la frecuencia de las causas posibles que se marcaron antes, en caso de que sea un estudio de causas hipotéticas.

Instrumentos para la recolección de datos

- Incluir los instrumentos para la recolección de datos que se desarrollaron para medir los indicadores anteriores.

Exhibición de los datos

- Presentar los resultados en cuadros adecuados, incluyendo la representación gráfica (histogramas, diagramas de barras o diagramas "Pareto").

Identificar causas de fondo

- Basándose en el análisis, el equipo debe indicar qué causa o causas quisieran abordar a fin de solucionar el problema y mejorar la calidad. Es importante acordarse de concentrarse en las "pocas causas fundamentales" del problema que los datos hayan revelado.

Formular una solución

- Formular una descripción clara y sencilla de las mejoras propuestas para el proceso. Resumir los pasos clave del plan para poner en ejecución la solución. Describir los métodos utilizados para diseñar y seleccionar una solución (opcional).

Resultados

- Mostrar los datos que demuestran los adelantos logrados en términos de mejoras en el proceso y los resultados deseados. Se pueden utilizar los mismos indicadores que se utilizaron para analizar el problema o un subconjunto de los mismos. Los resultados deben demostrar que la discrepancia demostrada en el enunciado del problema se ha reducido, logrando que el desempeño real se parezca más al desempeño deseado.
- Es conveniente que los resultados se midan nuevamente después de 6 meses o un año para asegurar que las mejoras se mantienen. Si se ha producido el problema nuevamente, el equipo debe estudiar la situación para comprender el motivo y se deben tomar otras medidas correctoras.

Anexo 2.2

Instrucciones para la realización de la técnica de grupo nominal

■ PREGUNTA: ¿Cuáles son las () oportunidades de mejora más importantes en tu ámbito de trabajo?

1. Generación de ideas

- Cada participante anota sus respuestas de forma individual y sin comentario alguno.

2. Enunciado de ideas

- Cada participante enuncia su primera opción y se efectúan tantas rondas como sean precisas hasta agotar las ideas anotadas.
- Todas las ideas deben recogerse en una pizarra o similar.
- Las ideas sólo se enuncian no permitiéndose aclarar, opinar o discutir sobre su contenido.
- Si durante esta fase se generan nuevas ideas, también se apuntarán.

3. Clarificación

- Se trata que los participantes asignen el mismo significado a los enunciados. Para ello, cada participante explicará a los demás el significado del enunciado que expuso. No se deberá opinar o discutir sobre el contenido de los enunciados.

4. Primera votación

- Se hace de forma individual y secreta, puntuando, en orden decreciente según su importancia.
- Después se recogen las puntuaciones y se anotan juntos a las opciones elegidas.

5. Discusión

- Se procede a discutir sobre los resultados obtenidos en la fase anterior, justificando y criticando los mismos de forma libre y ordenada. Se establecen los turnos de palabra fomentando la participación de todos los componentes del grupo.

6. Rondas de votación-discusión

- Se procede a sucesivas rondas hasta obtener las oportunidades de mejora prioritarias.

7. Informe del resultado

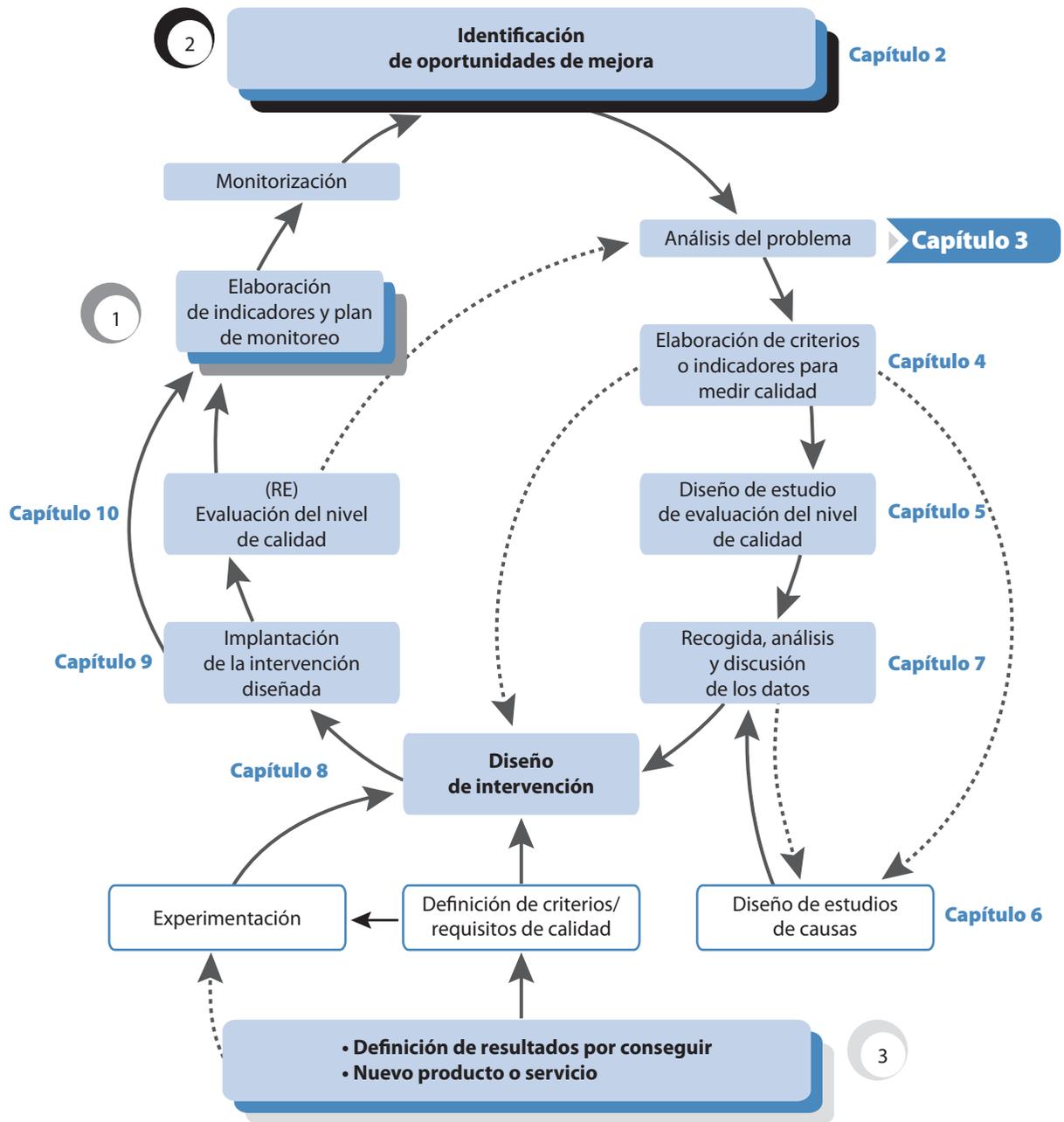
Capítulo 3

Métodos de análisis de los problemas de calidad

Contenido general y objetivos específicos

1. Introducción
2. ¿Por qué se realiza el análisis del problema de calidad u oportunidad de mejora?
3. Métodos de análisis de problemas
4. Análisis de problemas con métodos que no requieren datos
5. Herramientas de análisis que requieren datos sobre el problema

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Puntos de partida

Vía de acción normal

Vía contingente o alternativa

1 Monitorización

2 Ciclo de mejora

3 Diseño

Contenido general

Una vez delimitado el problema de calidad u oportunidad de mejora por evaluar (capítulo 2), este capítulo desarrolla la metodología para analizarlo con datos y sin ellos, y sienta las bases para las acciones o pasos siguientes. Se revisan diversos métodos útiles para analizar problemas de calidad y se explica en detalle y con ejemplos prácticos el uso para estos fines del diagrama de Pareto, histograma, flujograma y diagrama de causa-efecto. Estos últimos son los más útiles y utilizados para analizar problemas de forma grupal y cualitativa, por lo que se recomienda ponerlos en práctica frente a algún problema del entorno del propio lector.

Objetivos específicos

1. Describir los objetivos que se persiguen al analizar los problemas de calidad con cualquiera de los métodos de análisis que se utilice.
2. Identificar los principales tipos de problemas o situaciones para los que está indicado cada uno de los métodos.
3. Especificar las condiciones de aplicación de los métodos de análisis sin datos.
4. Especificar las limitaciones en la interpretación de los resultados de los análisis de problemas sin datos.
5. Definir el significado y distinguir los símbolos utilizados en un flujograma.
6. Analizar estructuradamente un flujograma.
7. Realizar un diagrama de causa-efecto para analizar problemas de calidad.
8. Analizar los resultados de un diagrama de causa-efecto.
9. Distinguir entre los tres principales tipos de causas que pueden ser identificadas al analizar los problemas de calidad.
10. Formular diversos cursos de acción sobre la base del análisis efectuado.
11. Decidir el curso o cursos de acción consiguientes, según los resultados del análisis efectuado.
12. Tener en cuenta las implicaciones metodológicas y los riesgos de los diversos cursos de acción.

“No podemos considerar nada verdad si no reconocemos claramente que lo es. Para conseguir esto, puede que sea necesario dividir un problema complejo en cuantas partes parciales sea posible. Entonces se puede empezar por las ideas más sencillas”. Descartes

1. Introducción

En el capítulo anterior habremos identificado una oportunidad de mejora. Ahora veremos que hay varias metodologías de utilidad para analizar las oportunidades identificadas, incluso sin tener datos. Un análisis correcto nos aclara el camino que habremos de seguir. Para decidir sobre las alternativas de acción que conduzcan a la mejora del probable problema de calidad identificado, debemos reflexionar acerca del grado de conocimiento que tenemos sobre el mismo (¿cómo es de importante? ¿cómo se puede corregir?) y realizar un análisis estructurado para identificar los factores implicados en él. Sobre esta base vamos a decidir qué hacer. En los capítulos siguientes avanzaremos por uno de los posibles cursos de acción, hasta completar el ciclo de mejora.

2. ¿Por qué se realiza el análisis del problema de calidad u oportunidad de mejora?

Una vez que hemos identificado la oportunidad de mejora, debemos preguntarnos de qué forma puede abordarse para conseguir su resolución. Sin embargo, es probable –de hecho es lo más frecuente– que en este segundo paso del ciclo de mejora tengamos un conocimiento insuficiente del problema y no estemos preparados para pasar al planteamiento de soluciones. Para ello, requerimos indagar sobre sus causas e incluso cuantificar hasta qué punto existe o no el problema. No es fácil cuantificar el problema ni identificar las medidas de intervención cuando éste se aprecia como un todo,

en conjunto y sin un conocimiento explícito y visualizado de sus implicaciones. Normalmente conviene que lo examinemos y lo dividamos en aspectos (subproblemas o causas) que sean lo suficientemente manejables para determinar, en cada una de ellas, el tipo de contribución al problema analizado y el curso de acción consiguiente que conviene emprender. En consecuencia, vamos a realizar el análisis estructurado de la oportunidad de mejora para conseguir los siguientes objetivos: dimensionar el problema; identificar las causas; clasificarlas y priorizarlas, y decidir cómo continuar en el camino hacia la mejora.

Dimensionar el problema. ¿Constituye la oportunidad de mejora un problema de calidad real?, ¿qué magnitud tiene el problema? La respuesta es más fácil si los métodos empleados en la identificación de la oportunidad de mejora han sido de tipo cuantitativo, como la monitorización, el análisis de perfiles o el informe del usuario, o bien si ya disponemos de datos relacionados con ese tema. Sin embargo, con el resto de los métodos intuimos una divergencia entre lo que se observa y lo que es razonable que exista (se ha sospechado un problema de calidad) pero no se ha objetivado todavía. En estos casos la reflexión sobre sus posibles causas y la frecuencia con la que se dan éstas puede confirmarnos intuitivamente la importancia del problema. Sin embargo, el dimensionado real no se conocerá (a menos que tengamos ya cuantificadas las causas), hasta que no se mida.

Identificar las causas. Consiste en descomponer el problema en unidades elementales que sean

En este capítulo se sientan las bases para el camino que habremos de seguir en la continuación del ciclo de mejora, tras haber identificado qué se quiere mejorar.

El análisis del problema de calidad u oportunidad de mejora se realiza para dimensionar el problema, reflexionar sobre sus causas y decidir qué hacer a continuación en el camino hacia la mejora.

susceptibles de medición y que a la vez se puedan traducir en medidas operativas que faciliten la resolución del problema. Por ejemplo, la existencia de hojas sin numerar puede ser una de las causas de que no se pueda revisar la historia clínica de los pacientes atendidos (que a su vez puede contribuir como causa al problema de calidad que estamos analizando). Además de ser medible, se puede convertir en una medida correctora, anulando, contrarrestando o modificando la propia causa, que en este caso sería simplemente la numeración de las hojas de la historia clínica.

Clasificación y priorización de las causas. Con el análisis estructurado, es posible identificar las principales subáreas problema (componentes principales del problema) y las causas del problema (componentes medibles y operativos para la intervención) a la vez que priorizar el orden de intervención para su corrección o estudio, según su importancia y el grado de conocimiento que tengamos sobre su frecuencia.

Decidir cómo continuar. Éste es el principal objetivo del análisis del problema u oportunidad de mejora. Como veremos más adelante, lo más probable es que nos encontremos con diversos tipos de causas frente a las cuales la acción más oportuna es también diferente. Resumidamente, habrá tres tipos de causas: a) causas sobre las que no podremos actuar fácilmente, típicamente causas estructurales en relación con la adquisición de recursos; b) causas sobre las que será posible la intervención directa, porque sabemos que influyen en el problema de calidad analizado y sabemos su grado de presencia en nuestro centro: también suelen ser causas estructurales y organizacionales pero a nuestro alcance; por ejemplo, la inexistencia de protocolización, la falta de numeración de las páginas en la historia, circuitos innecesariamente complicados o redundantes, etcétera, y c) causas que tendrán que ser investigadas y cuantificadas. Si resumimos aún más, los dos cursos de acción a los que nos va a guiar el análisis del problema son la intervención directa y el diseño de estudios para un conocimiento cuantificado de las causas sobre el que basar la intervención.

Todos estos objetivos pueden lograrse utilizando métodos muy diversos, cuya aplicabilidad depende del tipo de problema y de la existencia o inexistencia de datos sobre el mismo.

3. Métodos de análisis de problemas

A continuación, vamos a revisar los métodos que en general son de más utilidad, clasificados según precisen o no de la existencia previa de datos.

No hay reglas fijas para elegir un método de análisis determinado. Lo importante es que el método o métodos elegidos se adapten al problema y permitan alcanzar los objetivos del análisis. La única limitación radica en la disponibilidad o carencia de datos que cuantifiquen las subáreas o las causas del problema. En el cuadro 3.1 se resumen los métodos que vamos a revisar.

Cuando no tenemos datos, podemos analizar básicamente el problema con tres tipos de herramientas: el flujograma, el diagrama de causa-efecto y el análisis de campos de fuerza (ver algoritmo representado en la figura 3.1).

El flujograma nos será de gran utilidad en problemas relacionados con procesos que tienen numerosas actividades secuenciadas en el tiempo, como en el caso de los problemas de tipo organizacional, circuitos, coordinación de profesionales y niveles, entre otros (por ejemplo, la cita programada, la organización de médicos con enfermeros en el seguimiento de pacientes, el circuito de obtención de pruebas complementarias desde que se solicitan hasta

Cuadro 3.1
Herramientas de análisis de la oportunidad de mejora

1. Sin datos
 - 1.1. Diagrama de flujo
 - 1.2. Diagrama de causa-efecto
 - 1.3. Análisis de campos de fuerza
2. Con datos
 - 2.1. Histograma
 - 2.2. Estratificación
 - 2.3. Gráfico de Pareto
 - 2.4. Gráfico de control

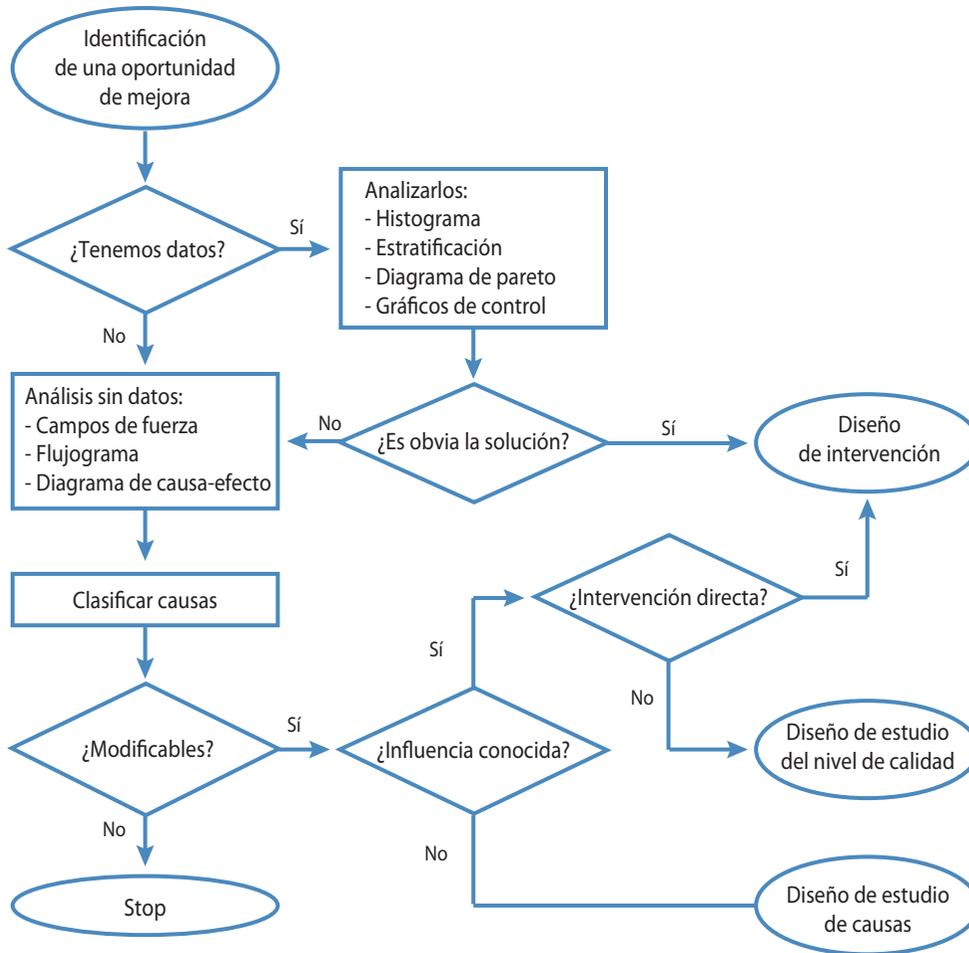


Figura 3.1. Algoritmo decisional del análisis de la oportunidad de mejora

que se obtienen, etc.); a su vez, el diagrama de causa-efecto tiene utilidad para todo tipo de problemas. El análisis de campos de fuerza, de utilidad como método de diseño de estrategias, puede ser también utilizado para identificar causas sobre las que intervenir o cuantificar. Por último, aunque cada una de estas herramientas tiene utilidad por sí sola, pueden también utilizarse combinadas: por ejemplo, un flujograma puede identificar dónde están los problemas de un circuito o proceso determinado, y se puede aplicar un diagrama de causa-efecto para analizar las causas.

Aunque no precisan de datos, la principal condición para el empleo de estos métodos es que se tenga experiencia sobre el problema analizado con el fin de aportar la información

necesaria para el análisis. Su empleo será adecuado, por tanto, en la exploración de oportunidades de mejora que nos sean conocidas de forma directa. Por otra parte, aunque en principio los podemos elaborar de forma individual, la información obtenida es mayor y, en consecuencia, el análisis es más exhaustivo si se hace en grupo; más aún si este grupo, conocedor del tema, es el que se va a encargarse de solucionar el problema, es decir, si es lo que se llama un grupo de mejora.

Si tenemos datos del problema, el análisis se puede realizar utilizando el histograma, la estratificación o el gráfico de Pareto, y, si tenemos una serie temporal de mediciones, los gráficos de control. Con estos métodos estaremos más seguros sobre qué hacer a continuación,

qué intervenciones son oportunas o hacia dónde dirigir estudios en más profundidad. Es obvio que su gran limitación se encuentra en la disponibilidad de los datos necesarios.

Una vez enunciados, vamos a ver en qué consisten estos métodos.

4. Análisis de problemas con métodos que no requieren datos

4.1. Flujograma

Con el diagrama de flujo o flujograma vamos a describir, de forma estructurada, la secuencia de actividades de un proceso determinado.

Es una de las herramientas más útiles para analizar problemas organizacionales o de procesos que incluyen numerosas tareas secuenciadas a lo largo del tiempo (por ejemplo, la organización de los distintos profesionales en el seguimiento de un paciente crónico, los circuitos de petición de pruebas complementarias, etc.). Con él simplificamos un proceso complejo dividiéndolo en actividades definidas, en las que sea más fácil apreciar en cuál de ellas o en qué paso entre ellas puede haber problema. De hecho, la percepción directa de procesos complejos sería análoga a captar, en una sola toma instantánea, un objeto en movimiento, con la consiguiente pérdida de información. Mediante el flujograma conseguiremos apreciar a cámara lenta este movimiento para identificar cada tarea del mismo y apreciar así las interconexiones y relaciones existentes.

Entre los principales componentes para plasmar en un flujograma podemos distinguir: a) una entrada o inicio; b) actividades concretas; c) situaciones de contingencia, donde se comprueba si ocurre o no determinada circunstancia decisiva y d) una salida. Estos elementos se representan mediante símbolos preestablecidos y su secuencia representa el proceso analizado (figura 3.2):

- Elipse: representa el inicio (lo que se requiere para comenzar el proceso) y el final del proceso analizado.
- Flecha: señala la dirección en la que fluye el proceso dentro de la secuencia de tareas.
- Rectángulo: representa una tarea o una acción determinada dentro del proceso. Normalmente se sigue un solo flujo de activi-

dad; de lo contrario, se requerirá intercalar un bloque de contingencia o decisión.

- Rombo: señala una situación de contingencia o comprobación que dividirá la dirección del flujo en dos o más opciones y en la que la conveniencia de elegir una dependerá de que se dé una circunstancia determinada. Por tanto, se sigue de tantos flujos como posibilidades tiene esta contingencia, aunque habitualmente suele ser dicotómica: sí o no.

Para elaborar el flujograma se pueden seguir los siguientes pasos:

1. Identificar las actividades o tareas (por ejemplo, mediante una lluvia de ideas o cualquier otro medio de reflexión grupal) y determinar quiénes son sus responsables respectivos.
2. Clasificar las tareas en orden cronológico y determinar los flujos del proceso (¿qué dirección sigue?).

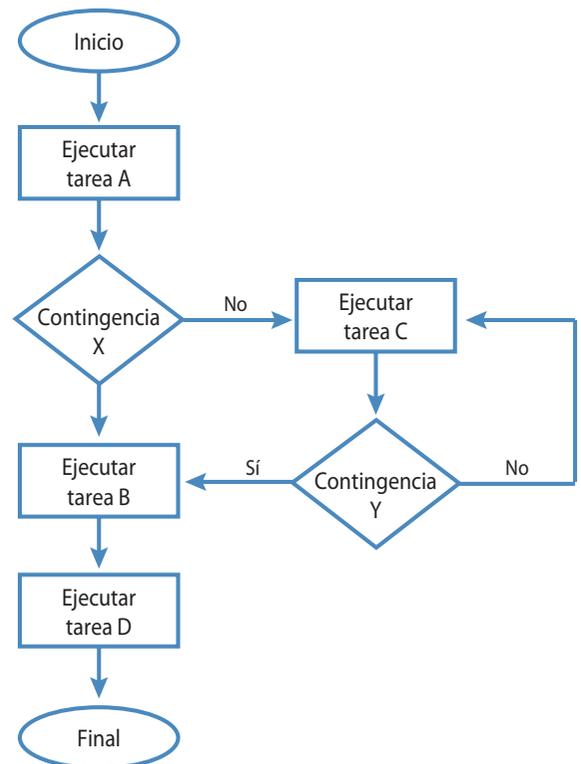


Figura 3.2. Representación gráfica de un proceso mediante el diagrama de flujo

El análisis del flujograma se realiza considerando una serie de aspectos concretos. Adicionalmente, puede utilizarse combinado con otros métodos de análisis para indagar las causas de los puntos señalados como críticos o problemáticos.

3. Identificar situaciones en las que una determinada tarea no se puede empezar porque la anterior no se ha concluido adecuadamente o no ha ocurrido algo que se supone debía haber ocurrido y, determinar qué actividad o contingencia es necesaria para que se resuelva esta situación y en qué orden del cronograma se colocaría.
4. Representar el flujograma y sustituir las situaciones en las que surge más de un flujo de una sola tarea por un nódulo de contingencia (con un rombo). Posteriormente, habrá que señalar de qué tipo de contingencia se trata (dentro del rombo) y cuáles son sus posibilidades al lado de sus flujos respectivos. En este punto se puede dar por terminado el flujograma como representación gráfica del proceso, aunque generalmente es posible continuar identificando actividades que probablemente no se hayan identificado en un principio, concentrando nuestro foco de atención entre dos actividades consecutivas y cuestionándonos si se requiere alguna actividad o alguna contingencia entre ellas.

Una vez que hemos elaborado el flujograma podemos plantear las siguientes cuestiones generales para analizar el problema de calidad:

1. Análisis del flujo general: ¿es lógico el flujo?, ¿hay diferencias entre el flujo ideal y el realizado de forma habitual?, ¿cuáles son las causas de estas diferencias?
2. Análisis de cada etapa del proceso: ¿de qué forma esta etapa puede no cumplir su función?, ¿cuáles son los problemas potenciales de calidad en esta etapa?, ¿cuáles son las posibles causas de estos problemas?
3. Análisis de los bucles o pasos hacia atrás: ¿cuáles son las causas que motivan un retraso en el desarrollo del proceso?

Esta herramienta se puede utilizar de forma combinada con el diagrama de causa-efecto cuando queremos indagar acerca de las causas de los puntos críticos del proceso.

La figura 3.3 contiene dos ejemplos (3.3.a y 3.3.b) de aplicación de flujograma, que pueden servir para ver más claramente en qué consiste. En el anexo 3.1 se da una explicación más detallada de la forma de analizar un flujograma.

4.2. Diagrama de causa-efecto*

El análisis de una oportunidad de mejora mediante el diagrama de causa-efecto es probablemente lo más frecuente e incluso lo aconsejable cuando no disponemos de datos del problema. En la mayoría de las situaciones en las que se identifica un problema es posible apreciar, de forma más o menos inmediata, un reducido número de causas. Pero, si no se examina con más profundidad, pueden pasar inadvertidos otros factores que, aunque no sean tan evidentes, sí son esenciales para comprender el origen del problema.

Con esta herramienta podemos descomponer una oportunidad de mejora en elementos identificables, medibles y operativos para la intervención (causas potenciales del problema). Una vez que se han identificado las causas, trasladaremos la reflexión sobre el grado de conocimiento del problema en conjunto, a la reflexión sobre el grado de conocimiento de sus causas potenciales. En este nivel de acercamiento estamos en mejores condiciones para conocer la frecuencia o cuantificar parte de sus causas. Cuando se tiene este conocimiento, el curso de acción más oportuno (diseño de intervención o estudio) aparece con mucha mayor claridad.

El diagrama de causa-efecto se conoce también como diagrama de Ishikawa (su inventor) o “diagrama de espina de pescado” (por el aspecto que tiene). Es probablemente una de las herramientas más sencillas y, sin embargo, más eficientes para determinar las causas de un problema de calidad o para detectar subáreas problemáticas. Representa la relación entre un problema de calidad determinado (efecto) y sus causas potenciales según se esquematiza en la figura 3.4. Para ello se exploran, de forma sistematizada y en sucesivas fases de profundización en el análisis, las distintas categorías o clases de causas y factores que contribuyen al problema, identificándolos y relacionándolos unos con otros.

Para elaborarlo se puede seguir una serie de pasos en los que siempre hay que pregun-

* Este apartado fue mejorado con las aportaciones de los asistentes al “Seminario de Formación de Tutores para la supervisión de la Formación a Distancia”, celebrado en Pamplona los días 26 y 27 de Junio de 2003.

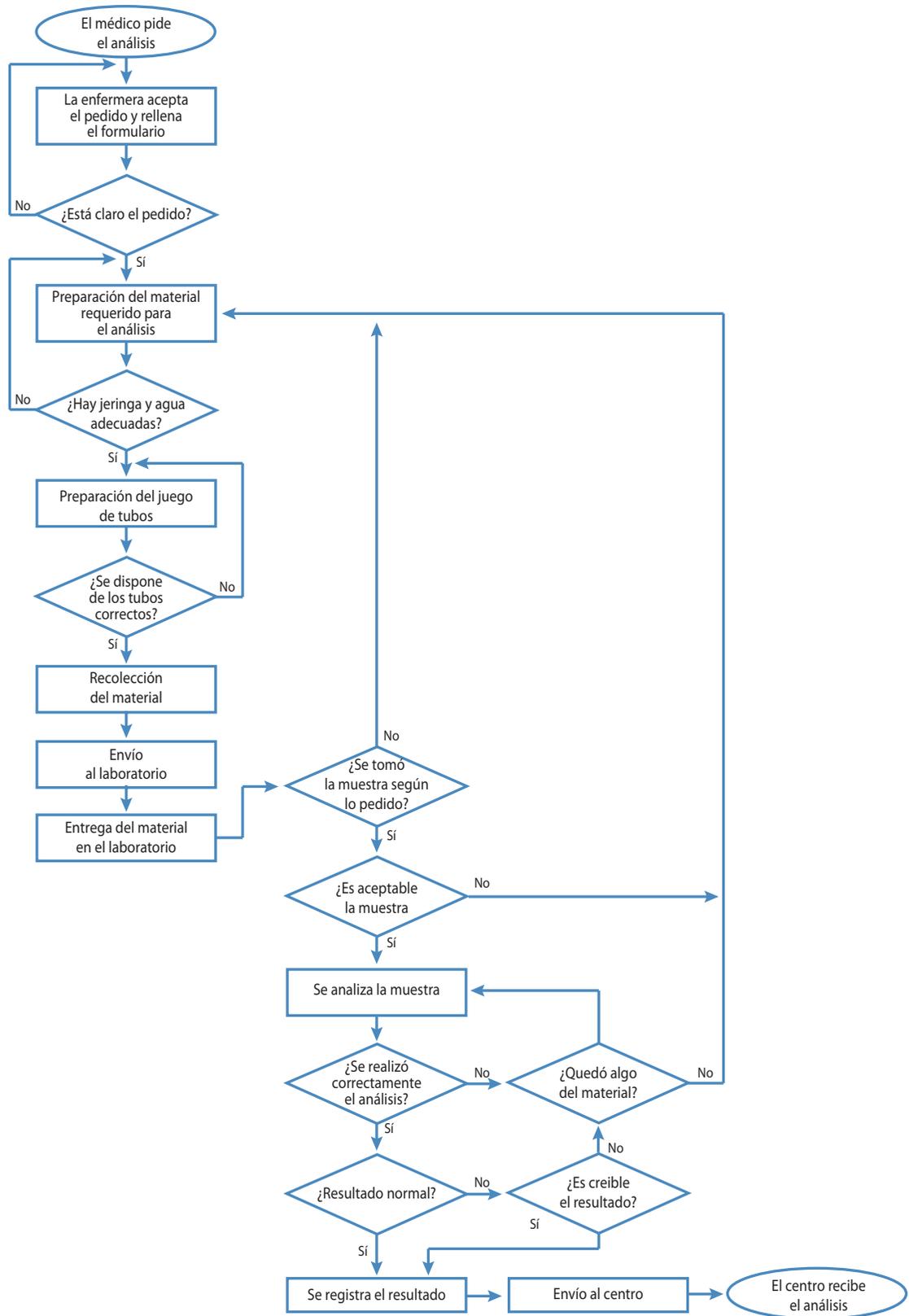


Figura 3.3.a. Flujoograma del circuito de petición y realización de análisis de petición y realización de análisis

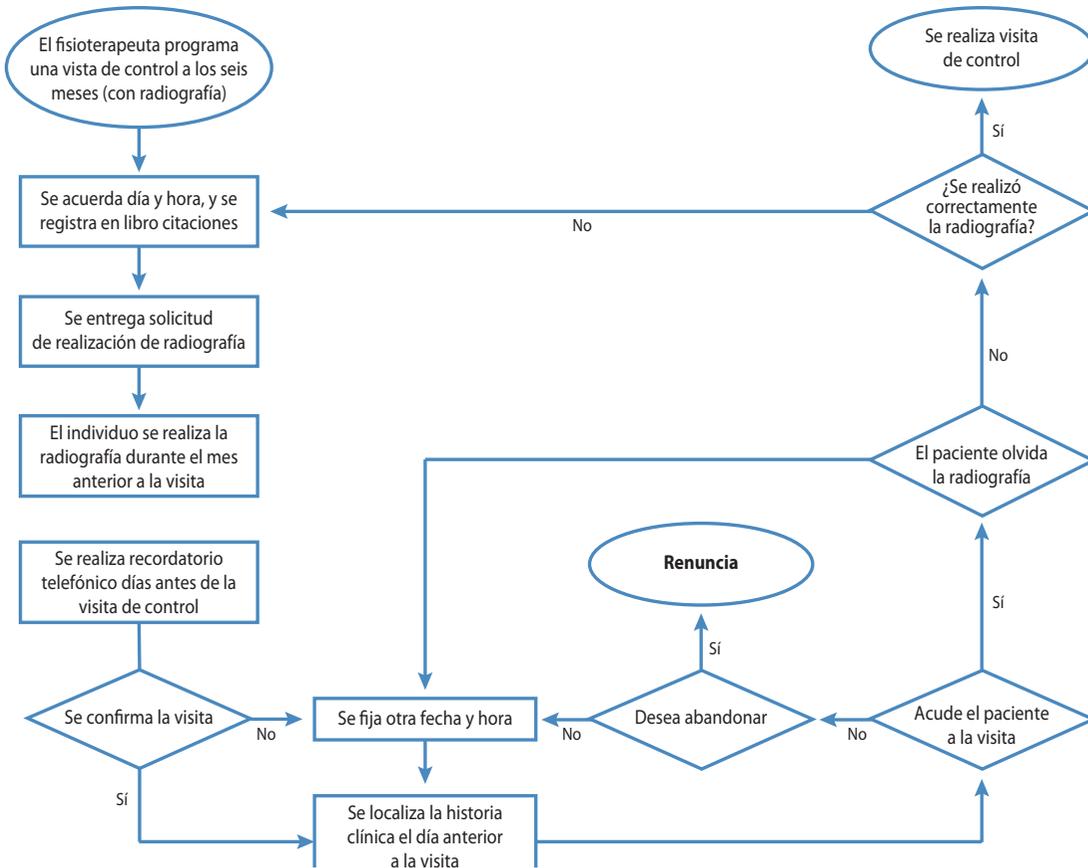


Figura 3.3.b. Flujo de seguimiento de la escoliosis

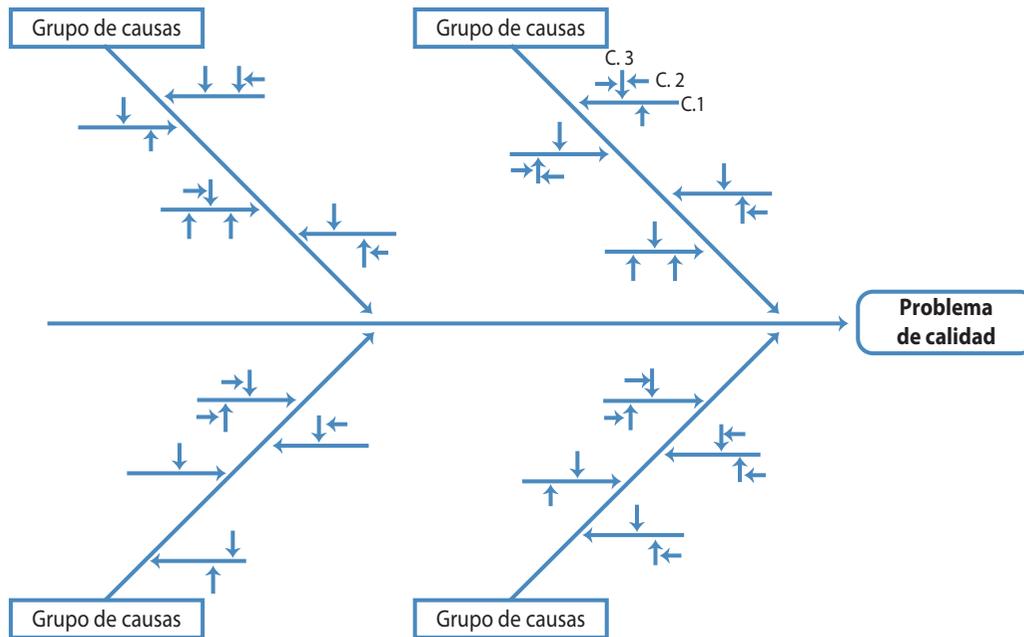


Figura 3.4. Diagrama de causa-efecto. Causas de primer orden (C.1), de segundo orden (C.2) y de tercer orden (C.3)

tarse: ¿Por qué ocurre la situación anterior? (problema, grupos de causas, causas de primer orden, causas de segundo orden...). Para ello se suelen utilizar técnicas que faciliten la creatividad como, por ejemplo, la lluvia de ideas (figura 3.5). Los pasos concretos para su elaboración son los siguientes:

1. *Anotación del problema.* Se traza una flecha centrada entre los márgenes superior e inferior y se coloca el problema de calidad o efecto en su parte derecha. El problema es la cabeza de la espina. Debe estar definido de forma escueta y clara (por ejemplo: "seguimiento irregular del paciente con escoliosis", "captación deficiente del hipertenso", "historia clínica no disponible en el momento de la consulta", "exceso de pruebas complementarias", etc.). Puede

ferior y se coloca el problema de calidad o efecto en su parte derecha. El problema es la cabeza de la espina. Debe estar definido de forma escueta y clara (por ejemplo: "seguimiento irregular del paciente con escoliosis", "captación deficiente del hipertenso", "historia clínica no disponible en el momento de la consulta", "exceso de pruebas complementarias", etc.). Puede

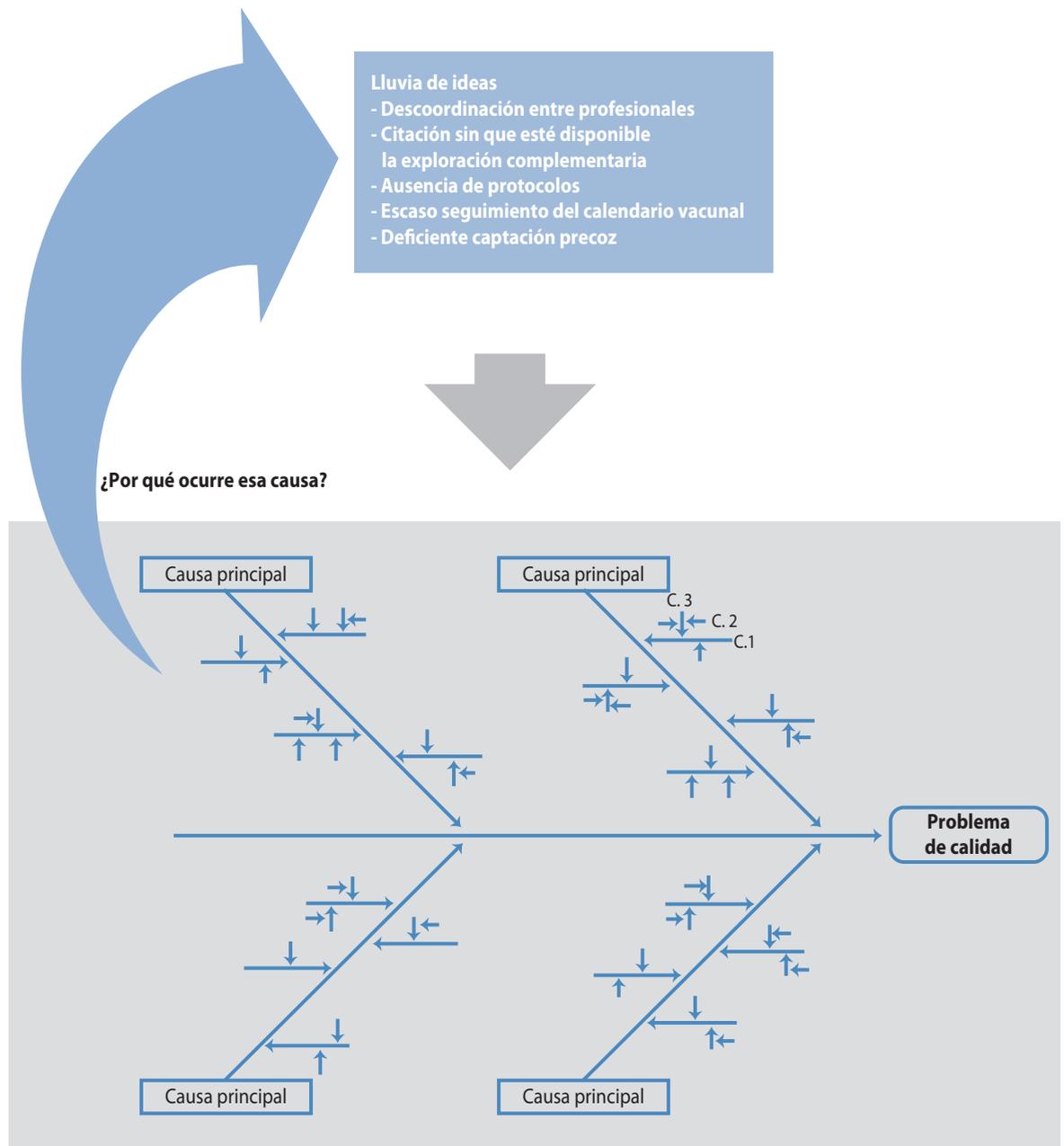


Figura 3.5. Identificación secuencial de las causas de un problema

incluso ser cuantificado (por ejemplo: “el 30% de los pacientes no vienen a las citas de control”, “el número de diabéticos en programa es menor a 50% de los esperados”, “el 10% de los pacientes tienen que ser vistos sin su historia”; “el 60% de los hipertensos no tienen controlada su cifra tensional tras seis meses de tratamiento”, etc.).

2. *Análisis de grupos de causas.* Se determinan las causas principales del problema agrupadas en categorías genéricas o en un número reducido de grupos de causas (de 4 a 6) como, por ejemplo, los métodos de trabajo, el personal, los recursos físicos (infraestructura), los pacientes, etc. Cada una de ellas se representa como una flecha paralela a la flecha central y dirigida en uno de sus extremos hacia la flecha correspondiente a la causa genérica. La determinación de este grupo reducido de grandes apartados puede hacerse con discusión simple o con lluvia de ideas. Incluso puede posponerse y realizarse después de hacer el listado de causas de primer orden.
3. *Análisis de causas de primer orden.* Mediante una lluvia de ideas se determinan las cau-

sas de cada una de las categorías genéricas. Consiste en cuestionarse, para cada una de las categorías genéricas, ¿por qué sucede ésta?; por ejemplo ¿por qué los métodos de trabajo son causa del problema? Se representan como una flecha paralela a la flecha central y dirigida en uno de sus extremos hacia la flecha correspondiente a la causa genérica. Si se comienza la indagación de causas a este nivel. Después de tener el listado, hay que agruparlas en 4 a 6 grupos genéricos, como vimos en el paso anterior.

4. *Análisis de causas de segundo, tercer y hasta n orden.* Se procede de forma análoga a la anterior. Para cada causa hay que plantearse, ¿por qué ocurre?, y de ésta, su vez, ¿por qué?, hasta llegar a la más elemental. Las respuestas se representan como ramificaciones de esa causa.

Un ejemplo simple de diagrama de causa-efecto, aplicado al problema *historia clínica no disponible en el momento de la consulta* puede verse en la figura 3.6.

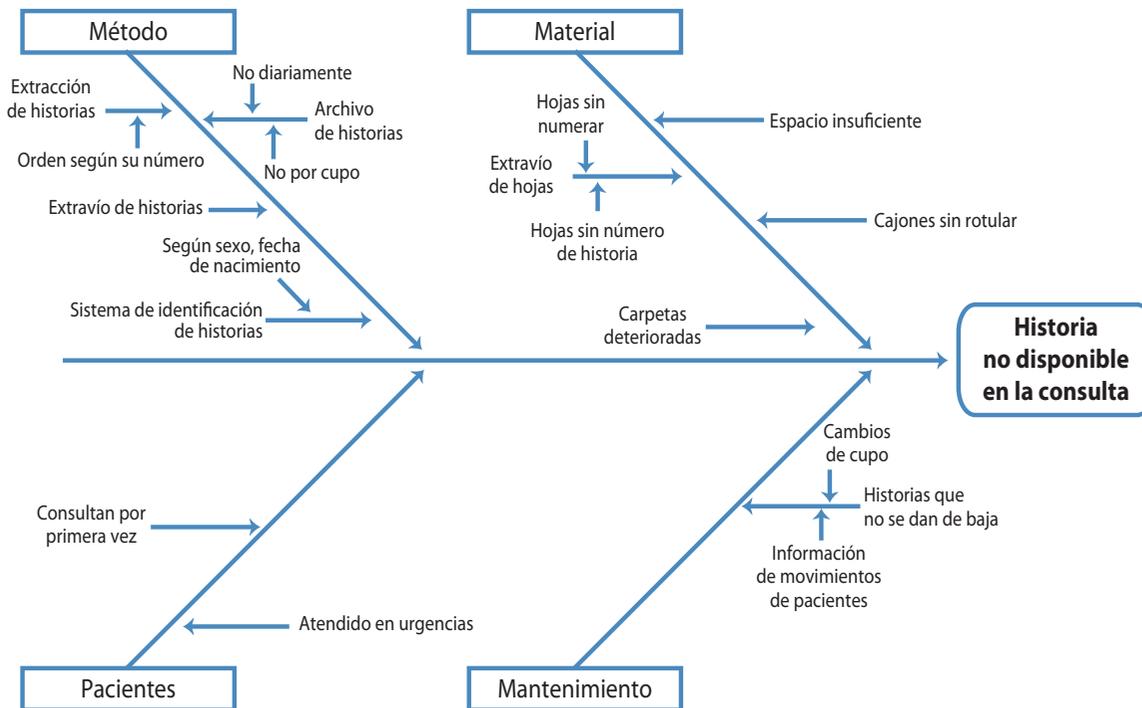


Figura 3.6. Diagrama de causa-efecto aplicado en el análisis de los “pacientes atendidos sin historia clínica” como oportunidad de mejora

Un punto siempre controvertido a la hora de hacer un diagrama de causa-efecto es hasta qué nivel de causas llegar, es decir, cuándo parar de interrogarse. Otro problema que suele ocurrir es encontrarse causas que encajan en más de un grupo, ¿dónde colocarlas? No hay una norma inflexible ni, por tanto, una respuesta única a estas cuestiones. En relación con el nivel al que hay que llegar en la explicitación de causas, en teoría tendría que ser hasta habernos quedado sin respuestas; sin embargo, en la práctica el proceso de indagación se detiene cuando se ve que esta causa ya es medible y se entiende bien cómo puede influir en el problema analizado. En relación con la ubicación de causas que pueden estar en más de un grupo, tampoco hay regla inflexible; lo importante es que estén, que se hayan identificado: si se desea, se pueden clasificar en más de una categoría, aunque lo lógico sería elegir aquel grupo al que contribuyan más, o al que sean más afines; por ejemplo, una causa como “no se indaga

adecuadamente en los hábitos alimenticios del paciente”, que puede contribuir a fracasos en el control del riesgo cardiovascular, puede ser incluida en el grupo de *métodos* o en el de *personal*, pero parece más claro que sea en el de *métodos*; otra cosa sería si la causa fuese “el personal desconoce la importancia de los factores dietéticos en el control del riesgo cardiovascular”.

Una vez que se han terminado de identificar las causas potenciales del problema de calidad, es posible clasificarlas, desde un punto de vista operativo, en las siguientes categorías (esquematisadas en la figura 3.7):

1. Causas estructurales inmodificables: son factores de la producción de servicios (recursos físicos y recursos humanos) sobre los que normalmente no podemos intervenir a corto plazo o que escapan de nuestra capacidad de intervención y que, por tanto, en principio, limitan y condicionan la calidad. Es bueno, en todo caso, conocerlos y tenerlos presentes.

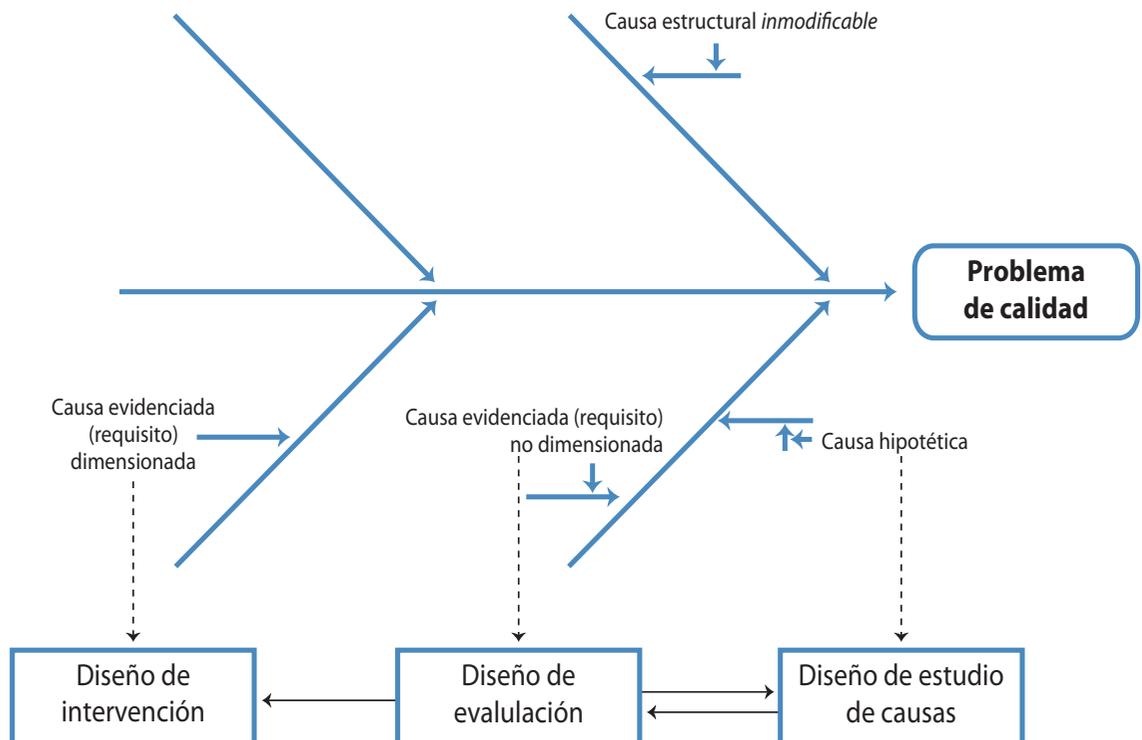


Figura 3.7. Clasificación de las causas de un problema de calidad y vías de acción consiguientes

2. Causas con repercusión conocida en el problema analizado (requisitos de calidad): son factores en los que hay evidencia previa de la existencia de un nexo causal con el problema de calidad. A su vez, se pueden distinguir dos tipos:
 - 2.1. Frecuencia conocida (cuantificadas): son requisitos de calidad normalmente estructurales y de organización de los cuales tenemos datos sobre la frecuencia con la que ocurren, que suele ser siempre o nunca, y sobre los que podemos intervenir directamente para mejorar la calidad. Por ejemplo, la ausencia de normas, protocolos, registros, existencia de circuitos redundantes, etcétera.
 - 2.2. Frecuencia desconocida (no cuantificadas): son requisitos de la calidad que no están cuantificados y que deberán medirse para determinar el nivel de calidad. Lo más frecuente es que sean actividades de proceso que no sabemos hasta qué punto se realizan. Se sigue, en consecuencia, con el diseño de evaluación, para saber su nivel de cumplimiento. En los servicios de salud, todos los requisitos de una buena actuación clínica están en este grupo.
3. Causas de repercusión hipotética en el problema analizado: son factores de los que intuimos una relación con el problema de calidad, pero sin contar con evidencia previa sobre esta relación. Su curso de acción consiguiente es el diseño de investigación de causas, para saber no sólo hasta qué punto están presentes o no, sino también hasta qué punto influyen en el problema de calidad analizado. Supongamos, por ejemplo, que nuestro problema es el seguimiento irregular del paciente hipertenso; para explicarlo, podemos pensar en el cumplimiento de los criterios establecidos en cuanto a periodicidad de controles y solicitud de exploraciones complementarias. Éstos serían requisitos que tendríamos que medir para saber el nivel de cumplimiento. Serían causas con repercusión previamente conocida pero frecuencia desconocida (la situación 2.2 explicada más

arriba). También podríamos pensar, como hipótesis, que los controles no se cumplen porque los análisis se pierden, o porque el paciente cambia las citas con demasiada frecuencia. Todas éstas son hipótesis plausibles pero que hay que medir e indagar si en realidad ocurren y si repercuten en los seguimientos irregulares (por ejemplo, un análisis que se pierda puede repetirse sin perjudicar en definitiva el seguimiento del paciente; lo mismo puede suceder con los controles de citas).

Los métodos de análisis de los problemas de calidad que no precisan datos son los que se utilizan con mayor frecuencia; sin embargo, son complementarios con los métodos que utilizan datos, en los casos en los que dispongamos de esta información.

5. Herramientas de análisis que requieren datos sobre el problema

5.1. Histograma

Cuando el problema por analizar está expresado en términos de un solo indicador o criterio de tipo cuantitativo (es decir, no mide presencia o ausencia de una cualidad, sino cantidad de una determinada variable, como puede ser tiempo de espera), su resumen estadístico simple (por ejemplo, la media) suele no ser suficiente y es conveniente realizar un análisis gráfico para saber qué hacer con el problema. Este análisis gráfico es el histograma tal como se estudia en los textos básicos de estadística, sólo que interpretado de una forma un tanto peculiar.

El histograma, tal como se aprende en las primeras lecciones de cualquier curso de estadística, es un gráfico de barras en el que se muestra la distribución de una determinada variable cuantitativa a partir de los datos de una tabla de frecuencias de los intervalos o valores sucesivos en que se ha dividido la variable que se ha medido; por ejemplo, distribuimos el tiempo de espera para la consulta en intervalos de cinco minutos y vemos cuántos han esperado de 1 a 5 minutos, de 6 a 10 minutos, etcétera. Lo que se consigue al representarlo gráficamente con el objetivo de

Una vez completo el diagrama, las causas expuestas se pueden clasificar en función de su traducción operativa para la continuación del ciclo de mejora.

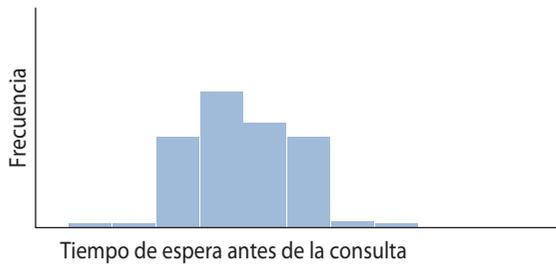
analizar el problema es, básicamente, ver el *tipo de variabilidad* que presenta, lo cual aclara la situación y nos ayuda a decidir cómo proceder a continuación. Por ejemplo, si tenemos medido el tiempo de espera para la consulta o los días que demora una determinada prueba analítica solicitada, no será suficiente sólo con saber el tiempo medio o la media de días porque desconocemos si estos tiempos medios representan una distribución homogénea o no. Tomemos el caso del tiempo de espera: al hacer el histograma podemos encontrar que hay poca variación y que la media representa bien lo que la gente espera; en este caso, si el tiempo de espera es elevado, el problema es generalizado y hay que diseñar un estudio de causas hipotéticas sin distinción de grupos de pacientes. Sin embargo, si nos encontramos que es bimodal (con dos picos) o francamente

sesgado a un lado, nos conviene identificar las características de los que esperan más (2º pico del histograma o cola extrema) porque son los que realmente representan el problema; o enfocar nuestro estudio en distinguir cuáles son las características diferenciales de los que esperan poco y los que esperan mucho (puesto que hemos visto que son dos grupos diferentes) y, en segundo lugar, si es necesario, cuáles son los factores adicionales que pueden asociarse a que se espere más. Este supuesto del tiempo de espera analizado con histograma está representado en la figura 3.8.

Para elaborar un histograma podemos seguir los siguientes pasos:

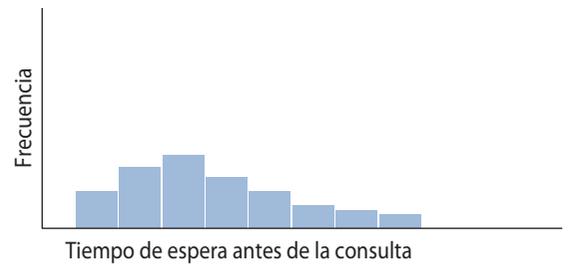
1. Recopilar los datos del problema o causa potencial del problema en estudio (por ejemplo, tiempo de espera antes de la consulta).

Al igual que para las demás herramientas de análisis estructurado de los problemas de calidad, la construcción de un histograma se realiza siguiendo una serie de indicaciones y pasos determinados.



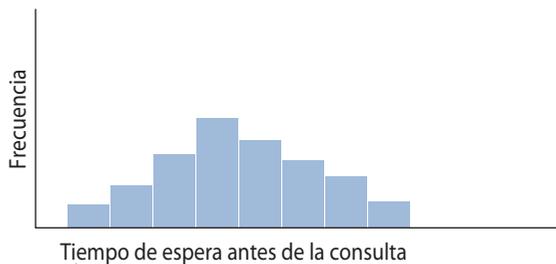
← Variabilidad pequeña →

Conclusión: problema general; estudio de causas sin distinción de grupos



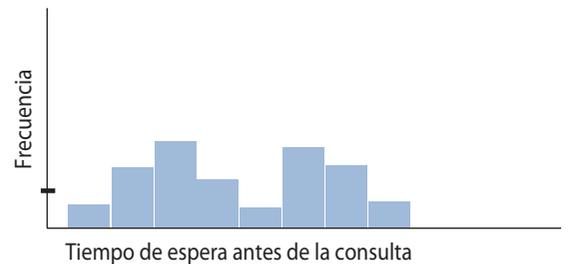
← Distribución sesgada a la izquierda →

Conclusión: problema no general; estudiar causas de espera en las que esperan más



← Variabilidad grande →

Conclusión: problema general; estudiar causas de variabilidad; ver sobre todo lo que puede ocurrir a las esperas por encima de la media



← Distribución bimodal →

Conclusión: problema no general; estudiar características diferenciadas de los dos grupos

Figura 3.8. Posibles histogramas del tiempo de espera antes de la consulta para decidir los estudios subsiguientes

2. Calcular el rango de la variable (restar el valor mínimo al valor máximo) y decidir sobre la amplitud del intervalo en que vamos a dividirlo. Supongamos que el rango resulta de 40', y que 5' nos parece un intervalo razonable.
3. Establecer el número de barras del histograma (debe resultar un mínimo de 6 y un máximo de 20) dividiendo el rango entre el intervalo establecido. En nuestro ejemplo tendríamos un histograma con ocho barras (40'/5').
4. Determinar la frecuencia para cada intervalo. Es decir, ver cuántos pacientes han esperado un tiempo incluido en cada uno de los intervalos.
5. Representar en un gráfico con dos ejes (un eje vertical "y" [frecuencia de cada intervalo] y uno horizontal "x" [los intervalos de la variable]) las barras correspondientes a cada intervalo.

Este gráfico reflejará el patrón de distribución de la variable analizada, así como la variabilidad de esta distribución y nos indicará el mejor camino para seguir (figura 3.8).

5.2. Estratificación

Una modalidad de analizar un problema de calidad consiste en presentar los datos de forma estratificada. En este caso, cuando tenemos los datos en conjunto de la oportunidad de mejora analizada (por ejemplo, tiempo transcurrido entre la recepción y la petición de pruebas de laboratorio) podemos saber el grado de contribución de una determinada causa potencial (por ejemplo, dos laboratorios distintos) recogiendo los datos de forma desglosada para cada uno de ellos y elaborando los dos histogramas. La contribución de esta causa potencial (laboratorio al que se solicitan las pruebas) puede apreciarse por el grado en que se diferencian ambos histogramas. Naturalmente, ante dos histogramas muy diferentes decidiremos estudiar o intervenir sobre el laboratorio más problemático.

La estratificación puede realizarse también cuando el problema sobre el que tenemos datos no es cuantitativo sino cualitativo. Supongamos, por ejemplo, que tuviésemos un servicio

de radiografía en nuestro centro y que el problema es que está colapsado, con listas de espera inaceptables. Un posible acercamiento, si tenemos datos de los motivos de prescripción de radiografías, sería estratificar por estos motivos de prescripción para ver cuál es más frecuente. Con este análisis por delante, lo lógico sería ver si están claras las indicaciones de radiografías para los motivos más frecuentes e intervenir directamente (para acordar suprimirlos) si aparecen motivos de solicitud de radiografía inadecuados. Sin la estratificación podríamos haber estado tentados a estudiar la adecuación de la prescripción radiográfica en general, en vez de centrarnos en las causas de prescripción más frecuentes e intervenir directamente sobre las inadecuadas. Situaciones como ésta, en la que la estratificación es de gran ayuda, se producen con frecuencia si el problema por analizar es de utilización de recursos (por ejemplo, derivación a especialista → estratificar por especialista o incluso por tipo de enfermedad; utilización de analítica → estratificación por tipo de análisis; prescripción de antibióticos → estratificación por patologías o tipo de antibiótico, etc.). Como veremos más adelante –capítulo 7–, la estratificación es uno de los métodos de gran utilidad en el análisis de los datos de un estudio de causas hipotéticas.

5.3. El diagrama de Pareto

Este gráfico se estudiará con detalle en el capítulo 7 como forma de análisis y presentación gráfica de los resultados de una evaluación. En general lo podemos aplicar para determinar en qué grado contribuyen las distintas causas o estratos de un problema en el mismo. En ambos casos lo que tenemos es un cierto número de categorías o de grupos y la frecuencia de su presencia o ausencia (variable cualitativa). Una vez medida la frecuencia con que se produce cada categoría, podemos ordenarlas de mayor a menor frecuencia y focalizar nuestra atención en las categorías más frecuentes.

El diagrama de Pareto consiste básicamente en un diagrama de barras, en el que cada barra es una de las categorías, criterios o estratos considerados, y en el que *las barras se ordenan de mayor a menor frecuencia*. De esta forma se hace gráficamente evidente qué es prioritario

La estratificación ayuda a entender la influencia de un determinado factor en la magnitud del problema y ayuda a priorizar los estudios de causas por realizar.

Un diagrama de Pareto sirve para priorizar las categorías o causas principales sobre las que dirigir nuestra atención. En el capítulo 7 se describe en detalle su construcción y utilización para analizar los datos de una evaluación y orientar en el diseño de las intervenciones.

analizar. El ejemplo de la estratificación de las radiografías por tipo de patología o cualquier otra estratificación por variables categóricas (variables con varias categorías o posibilidades excluyentes) pueden ser presentados gráficamente con diagrama de Pareto.

Hay que advertir, no obstante, que el análisis con diagrama de Pareto que veremos en el capítulo 7 se construye de una forma diferente y particular, y tiene mayor utilidad para decidir sobre diseño de intervención, que sobre estudios sucesivos.

La figura 3.9 es un diagrama de Pareto, hecho para decidir por dónde empezar a analizar la problemática de la derivación a especialistas.

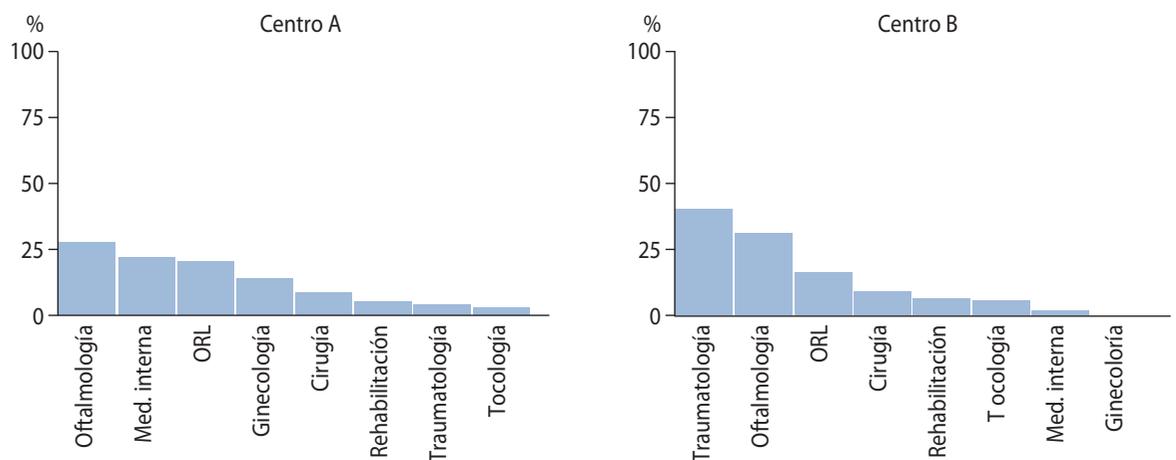
5.4. Gráficos de control estadístico

El control estadístico (gráfico) de la calidad es el método de elección para monitorizar indicadores que se miden de forma continua o muy frecuente (como mínimo cada mes). A esto hay que añadir su utilidad como herramienta de análisis de los procesos, lo que permite detectar no sólo si hay problema (algo común a todos los métodos de monitorización), sino también qué tipo de problema es (esporádico o sistemático). Al mismo tiempo, los gráficos

de control estadístico son sensibles a cualquier cambio introducido en el proceso que monitorizan, por lo que pueden utilizarse también para evidenciar los efectos de las intervenciones que se implementan para mejorar las situaciones que se han identificado como mejorables o problemáticas. Los gráficos de control estadístico son, pues, instrumentos muy versátiles, con múltiples aplicaciones.

El análisis básico que se realiza con el control estadístico consiste en determinar si los valores que se obtienen del indicador que se mide están dentro de los esperados, dada la variabilidad, o si son estadísticamente incompatibles con la situación estable de partida. Sin embargo, este análisis es visual, gráfico y no entraña, una vez construida la plantilla gráfica, ningún cálculo adicional a la propia medición del indicador. El objetivo inicial principal del análisis gráfico es caracterizar la variabilidad del indicador que se mide, dado que su mejora y control han de empezar por la comprensión de su variabilidad. Una vez conocida y controlada la variabilidad, el proceso es predecible y cualquier desviación que ocurriese será fácilmente detectable.

La gran utilidad potencial del control gráfico de la calidad contrasta con lo lenta que



Conclusión: en el centro A se priorizará el estudio de las derivaciones de oftalmología y medicina interna, mientras que en el centro B se priorizará el estudio de las derivaciones de traumatología y oftalmología

Figura 3.9. Pareto de derivación a especialistas

está resultando su utilización rutinaria en los servicios de salud. Probablemente la principal causa sea que se precisan series temporales con mediciones muy frecuentes para que puedan ser aplicados. Ésta es una limitación inexistente cuando hablamos de sistemas de monitorización frecuente de indicadores. Los gráficos de control estadístico se revisan en detalle en el documento *Métodos y herramientas para la monitorización de la calidad Manual* (volumen II de esta colección).

5.5. Análisis de causas raíz (*root cause analysis*)

Un tipo particular de análisis de causas es el conocido como análisis de causas raíz (*root cause analysis*, representado con las siglas RCA o RoCA, según versiones), que puede ser definido como el proceso de averiguar y eliminar la causa (o causas) últimas de un problema para evitar que vuelva a ocurrir. El método tiene su origen en el ámbito industrial con el objetivo de analizar las causas de los accidentes o sucesos adversos infrecuentes pero importantes, de forma que pudiera prevenirse su repetición. En esencia, se trata de una investigación estructurada a fondo para averiguar cómo podría haberse evitado ese accidente. Sin embargo, el campo potencial de aplicación se ha ampliado hasta incluir problemas recurrentes y frecuentes, así como variabilidad inadecuada en los resultados que se obtienen.

La incorporación de este método al campo sanitario es relativamente reciente, dentro del énfasis en los problemas de seguridad del paciente y particularmente en la vigilancia de sucesos centinela y el apoyo metodológico para su análisis. Ante la aparición de un suceso centinela, la JACO incluye como obligatorio el RCA así como los planes de acción consiguientes, en los estándares para acreditar. Un *suceso centinela* se define como “un suceso que ocurre de forma inesperada, implicando muerte o lesiones serias de tipo físico o psicológico, o un aumento del riesgo para ello en lo sucesivo”. Tanto el análisis de causas como el plan de acción debe realizarse dentro de los 45 días posteriores al suceso centinela y ser enviado a la JCAHO, que lo revisará y determinará si

es aceptable, con base en una serie de criterios que incluyen la búsqueda de causas relacionadas con el sistema y los procesos –no con las personas– y la credibilidad del plan de acción enviado.

El RCA a su vez incluye varias metodologías concretas, entre ellas los diagramas de causa-efecto y otras como el análisis de cambios y la comparación de procesos con malos resultados con procesos con buenos resultados, el análisis de barreras (evaluación de la adecuación de las barreras existentes para evitar el suceso adverso), el análisis secuencial de sucesos y factores concretos, etcétera.

Las causas pueden ser de varios tipos, y ser agrupadas como en el diagrama de Ishikawa, pero es interesante el énfasis en las causas atribuibles al sistema. Así, por ejemplo, se afirma como una de las reglas para realizar el RCA que “cada error humano ha de tener una causa que lo precede”. Descubrir que hubo un error humano no ayuda en nada a la prevención del problema. Es la causa del error, generalmente relacionada con el sistema o el diseño del proceso encausado, no el error en sí mismo, lo que puede conducirnos a estrategias de prevención productivas. El desarrollo exhaustivo de esta metodología, así como su práctica, escapa de los objetivos de este manual. No obstante, quienes estén interesados pueden revisar las publicaciones *Root cause analysis guidance document*, publicado en 1992 por la Oficina de Seguridad Nuclear, Política y Estándares del Departamento de Energía de los Estados Unidos y el Manual de la JCAHO *Root Cause analysis in health care: tools and techniques*, ambas en la bibliografía de este capítulo.

Con este capítulo estaremos en condiciones de analizar el problema para dimensionarlo, conocer sus causas, priorizar sus componentes principales y, en definitiva, decidir cómo continuar el ciclo de mejora. De todos los métodos expuestos se aconseja dominar la práctica de, al menos, el diagrama de causa-efecto sobre la calidad deficiente (o mejorable) y del flujograma, si es un problema de tipo organizacional o implica una secuencia de actividades.

BIBLIOGRAFÍA

- Banks NJ. Constructing algorithm flowcharts for clinical performance measurement. *Int J Qual Health Care* 1996; 8(4):395-400.
- Ishikawa K. Introducción al control de calidad [cap. 4]. Madrid: Diaz de Santos, 1994: 221-310.
- JCAHO. Root cause analysis in health care: tools and techniques, 4th ed. Oakbrook Terrace: Joint Commission Resources, 2010.
- Leebov W, Ersoz CJ. The health care manager's guide to continuous quality improvement [ch. 11]. AHA 1989:143-187.
- Manual de Herramientas básicas para el análisis de datos. Methuen: GOAL/QPC, 1990: 9-13; 17-23; 25-29; 36-43.
- United States Department of Energy. Root cause analysis guidance document. Washington DC: DOE, 1992. Disponible en: <http://www.nmenv.state.nm.us/aqb/documents/DOERootCauseAnalysisGuidanceManual.pdf> (consultado: septiembre de 2015).

Capítulo 4

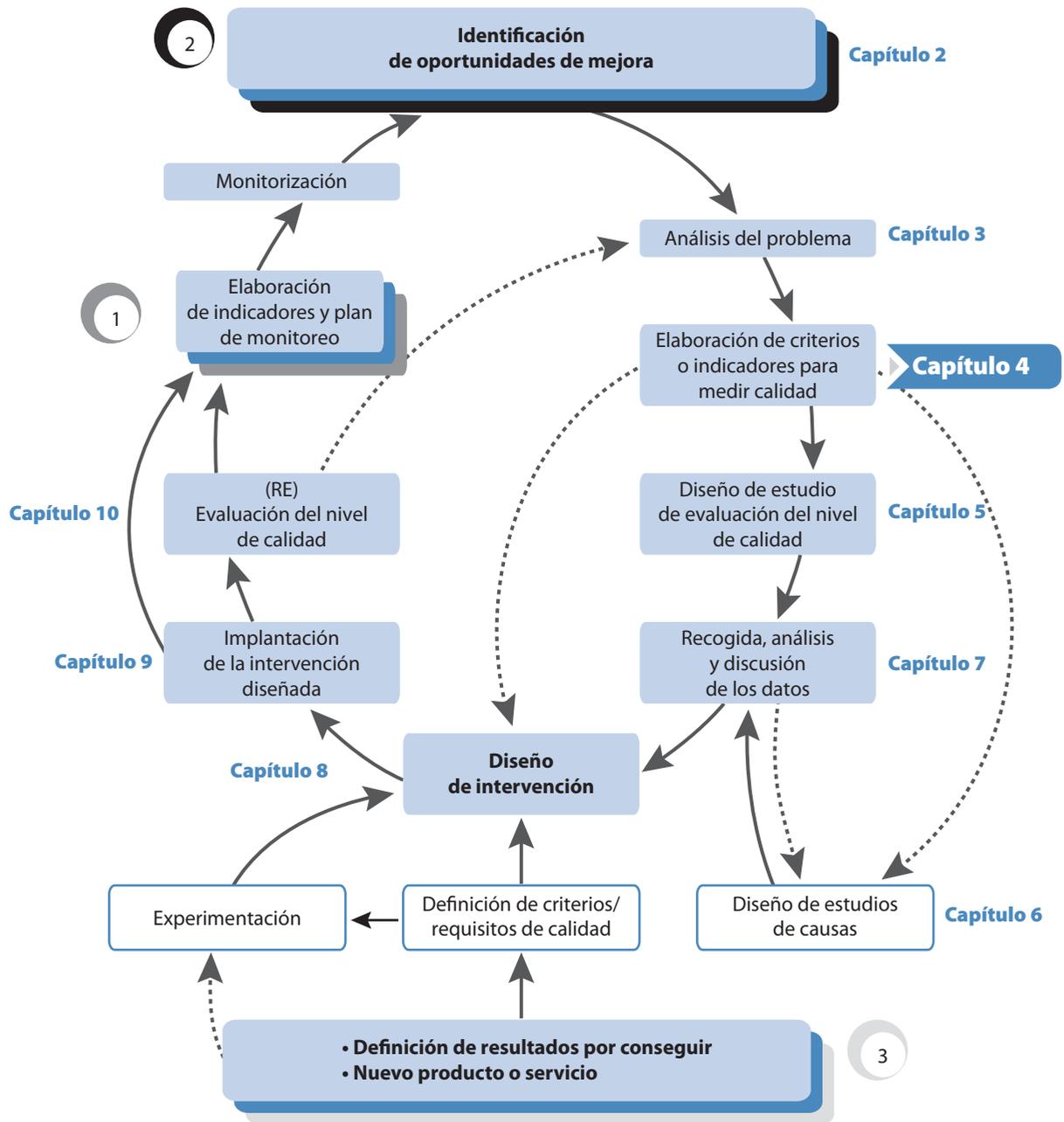
La construcción de criterios para evaluar la calidad

Contenido general y objetivos específicos

1. Introducción
2. ¿Qué es un criterio? Concepto de criterio para medir calidad
3. Tipos de criterios
4. ¿Cómo se construye un criterio?
5. Excepciones y aclaraciones a los criterios
6. Validez de los criterios

Anexo 4.1. Trabajo práctico recomendado: la construcción de criterios para evaluar la calidad. Análisis de validez

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Puntos de partida

Vía de acción normal

Vía contingente o alternativa

1 Monitorización

2 Ciclo de mejora

3 Diseño

Contenido general

El siguiente paso en el ciclo de mejora, sea cuál sea el curso de acción que decidamos tras analizar el problema (capítulo 3), es la construcción de los criterios, indicadores, herramientas o referentes con los que vamos a medir la calidad en la situación de partida, y que luego emplearemos para documentar la mejora conseguida. Además de otras consideraciones, los criterios o indicadores a emplear deben ser herramientas válidas y fiables, para lo cual es preciso abordar su construcción con un rigor metodológico que se detalla en este capítulo.

Objetivos específicos

Este capítulo está enfocado a:

1. Desarrollar criterios o requisitos para medir la calidad de un determinado producto o servicio, o de un aspecto del mismo
2. Distinguir los criterios aislados, transversales, de los secuenciales
3. Distinguir los criterios empíricos de los normativos
4. Discutir las ventajas y las desventajas de la ponderación de los criterios
5. Establecer mecanismos para asegurarse la validez de los criterios a utilizar
6. Describir los enunciados de los criterios acompañados de excepciones y aclaraciones de los términos empleados
7. Explicar el significado de las excepciones al criterio a la hora de medir la calidad
8. Determinar si las excepciones han de ser tratadas aparte o conjuntamente con los casos normales
9. Estar alerta de la posibilidad de que las excepciones sean un factor de confusión si se tratan conjuntamente con los casos normales

La construcción de criterios para evaluar la calidad

“Todo tiene que medirse y lo que no se puede medir, hay que hacerlo medible” (Galileo).

1. Introducción

Paso a paso hemos avanzado por el ciclo de mejora. Después de la identificación del problema de calidad (capítulo 2) y el análisis del problema seleccionado (capítulo 3), en este capítulo llegamos a una parte crucial: la construcción de instrumentos, requisitos, criterios o indicadores, con los cuales medir la calidad.

Vamos a ver que hay diferentes tipos de criterios con implicaciones metodológicas peculiares para su medición. Propondremos realizar las evaluaciones preferentemente con criterios explícitos y normativos y estableceremos las precauciones a tomar para asegurarnos de que tengan validez y fiabilidad.

2. ¿Qué es un criterio? Concepto de criterio para medir calidad

Los criterios son las herramientas o referentes con los cuales se mide la calidad; así como aspectos esenciales y relevantes de la atención sanitaria en torno al problema o área que queremos evaluar. Ese aspecto relevante de la atención sanitaria será medido para conocer su grado de cumplimiento; de forma que su ausencia, presencia o grado de cumplimiento, nos está indicando, a su vez, la presencia o los diferentes niveles de actuación correcta, es decir, diferentes niveles de calidad.

Un criterio se deriva de la reflexión sobre cómo han de ser los requisitos de los distintos servicios para que sean acordes con las necesidades y expectativas de quienes los van a recibir. En este sentido, es a la vez un objetivo de

calidad que queremos que se cumpla. De ahí la tremenda importancia que tiene el que sean relevantes, realistas y que puedan ser aceptados sin controversia por todos los profesionales implicados en su consecución. Una correcta elaboración y selección de criterios es probablemente, junto a una priorización adecuada del tema a evaluar, el principal factor determinante del éxito o fracaso de los ciclos de mejora.

Por ejemplo, supongamos que nos preocupe el nivel de calidad alcanzado en el programa de atención al cáncer en la mujer. No habría que discutir mucho para acordar que uno de los aspectos más relevantes para una atención correcta, de calidad, es la detección precoz. Así, de una manera general, la existencia o no de un programa de *screening* para el cáncer femenino es una medida de calidad. Sin embargo, un criterio estructural como es la existencia de programa para la detección precoz rara vez es suficiente. Lo habitual es que haya que especificar además cómo se ha de llevar a cabo el proceso, cuál es la traducción operativa de la satisfacción de las necesidades de salud que queremos cubrir. Siguiendo con nuestro ejemplo, la detección precoz ha de consistir en la práctica de citologías vaginales con una pauta determinada, y la realización también explícitamente pautada de exploraciones mamarias. El criterio o herramienta de medida sería, en este caso, el cumplimiento o incumplimiento de las pautas de actuación acordadas como concretas. De igual manera podríamos ir reflexionando sobre cuáles serían los requisitos para que no hubiera problemas de accesibilidad, atención centrada en el paciente, etcéte-

ra. La figura 4.1 contiene el resumen de este esquema básico de razonamiento para definir criterios de calidad.

Es adecuado recordar que los términos *criterio*, *indicador*, *requisito*, *norma*, *estándar*, conceptualmente emparentados, no son siempre utilizados de la misma manera en la literatura sobre calidad asistencial, más bien son utilizados de forma intercambiable y a veces confusa. La distinción terminológica tiene, sin embargo, una traducción práctica generalmente irrelevante. Lo importante para diseñar una evaluación no es tanto el nombre que le demos a la herramienta de medida, sino la secuencia consciente del razonamiento que nos lleva a definirla y a tener claro el concepto de lo que vamos a medir.

2.1. Características de los criterios

Con cualquiera de los servicios o aspectos de los mismos que consideremos como problema de calidad y queramos evaluar su nivel, el listado de criterios que podríamos construir es normalmente grande. Es conveniente, sin embargo, priorizar y prestar atención a un cierto número de características esenciales que distinguen los criterios buenos de los no tan buenos. Entre estas características que deben cumplir los criterios (cuadro 4.1), ya hemos hablado de la relevancia, una característica inherente al concepto mismo de criterio que hace referencia a su importancia y a que sea un aspecto esencial del problema a evaluar. También hemos hablado de la conveniencia de que

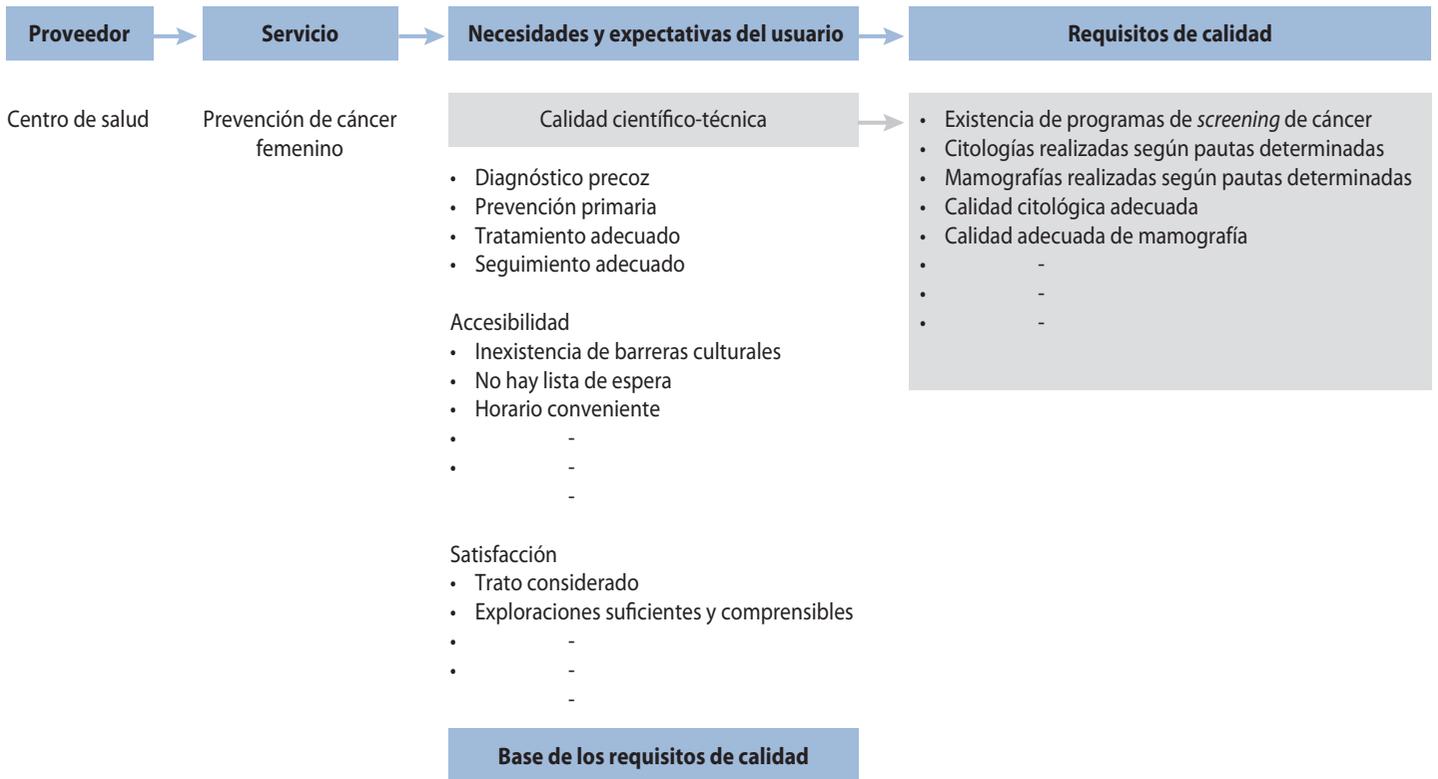


Figura 4.1. Ejemplo de aplicación del esquema de razonamiento básico para definir criterios para evaluar calidad

Cuadro 4.1
Características de los criterios para evaluar calidad

● Relevantes	● Medibles
● Realistas	● Fiables
● Aceptables	● Pocos
● Válidos	

dichos criterios sean realistas y aceptables, en cuanto a que sea algo que el centro sanitario y sus pacientes se pueden permitir (realistas) y que los profesionales y la bibliografía los sustente como científicamente comprobados en cuanto a niveles de calidad (aceptables y válidos). Hay que añadir a los requisitos que deben cumplir un criterio el que sea medible, se debe de poder cuantificar el nivel de calidad. También es aconsejable que sean pocos criterios, los menos posibles; algunos autores dan menos de 10 como referencia útil. Otra característica fundamental de los criterios, que veremos más a detalle en el siguiente capítulo, es que sean fiables. Fiabilidad y validez van a ser dos características de los criterios a los que vamos a prestar un interés especial.

2.2. Estándares de calidad

Si existen datos de estudios previos sobre los criterios de un determinado problema objeto de evaluación, estos nos permitirán establecer unos estándares o niveles óptimos de cumplimiento del criterio, los cuales se fijarán antes de iniciar la evaluación, de acuerdo con la situación asistencial concreta del ámbito donde se realice. Podemos definir *estándar* como el nivel de cumplimiento patrón que separa lo aceptable de lo no aceptable, es decir, un referente para enjuiciar el grado de aceptabilidad de la calidad encontrada. Es un límite que establecemos como nivel exigible de cumplimiento del criterio. En algunos casos vamos a exigir un 100% como estándar, y en otros (como generalmente ocurre en lo relativo a criterios de resultado), puede que consideremos estándares más bajos como aceptables. Aunque es aconsejable establecer estos estándares antes de evaluar, tal como hemos señalado, tampoco

es incorrecto establecerlos después de una primera evaluación.

En algunos ámbitos, el concepto de estándar no es diferente al de criterio. Por ejemplo, al referirse a los requisitos que se deben exigir y que se va a comprobar si están presentes al realizar una auditoría para acreditar o certificar, se emplea con frecuencia el término *estándar*.

2.3. Indicadores de calidad

Un tercer tipo de referente es aquel que vamos a emplear no tanto para evaluar el nivel de calidad, sino para concluir si hay o no un problema de calidad que evaluar. Es decir, una herramienta para hacer cribaje de problema de calidad más que un diagnóstico del nivel de calidad. A este referente le llamamos *indicador*, y lo vamos a utilizar en la monitorización de la calidad. Un indicador en realidad es un criterio, sólo que especial, porque tiene que representar o resumir al máximo la calidad de lo que queremos monitorizar. En este sentido podemos decir que todos los indicadores son criterios, aunque no todos los criterios son buenos indicadores. Por ejemplo, para evaluar la calidad del servicio de hipertensión podemos utilizar toda una serie de criterios en relación a todos los pasos del proceso asistencial; pero si lo que queremos es controlar periódicamente que el servicio se mantenga a un nivel aceptable, vamos a elegir solamente unos pocos criterios; cuanto menos mejor, pero que representen adecuadamente la calidad del servicio; a estos pocos criterios/resumen, son a lo que llamamos indicadores. Por ejemplo, la realización o no de los controles analíticos en el hipertenso en programa puede ser un buen criterio de calidad (a medir dentro de un listado más amplio) pero no un buen indicador porque resume mal la existencia o no de problemas de calidad en el servicio en su conjunto, mientras que la proporción de hipertensos con cifras normalizadas sí podría ser un buen indicador a monitorizar, debido a su relevancia y porque resume en sí mismo el efecto del cumplimiento de otros muchos criterios. El nivel de cumplimiento del indicador, que separa la existencia o no de problema de calidad, se suele conocer como *umbral* del indicador; un concepto paralelo a lo que hemos llamado estándar en

Los criterios han de ser relevantes, realistas, aceptables, medibles, pocos, válidos y fiables.

El estándar (nivel de cumplimiento patrón) de los criterios puede fijarse antes o después de evaluar.

relación a los criterios. El término indicador es muy frecuente encontrarlo como sinónimo de criterio, pero parece conveniente distinguir los conceptos de los dos tipos de referente que representan (uno para evaluar niveles de calidad: criterio; y otro para hacer screening de problemas de calidad: indicador), sea con esta o con otra terminología.

3. Tipos de criterios

Al hablar de criterios de calidad podemos encontrar diferentes tipos (cuadro 4.2):

- **Criterios implícitos:** son aquellos en los que no se especifica cuáles son los criterios de calidad que se valoran, se perfilan en función de los conocimientos del sujeto que realiza la evaluación. No es recomendable utilizar este tipo de criterio de forma exclusiva. Pueden emplearse más como mecanismo de detección de problemas o de segunda revisión de los casos clasificados de calidad deficiente. Ejemplo de criterio implícito: evaluar si el diagnóstico es correcto sin que se especifique cuál es el concepto de buen diagnóstico que emplea el evaluador.

Cuadro 4.2 Tipos de criterios
Implícitos: no descritos
Explícitos: descritos
Empíricos: de base estadística (lo que se hace)
Normativos: de base científica (lo que se debe hacer)
Secuenciales: el cumplimiento de uno se realiza en la población que cumple el anterior
Transversales: se valoran independientemente
Isovalentes: todos importan igual
Ponderados: a unos se les da más importancia que a otros
Genéricos: aplican a varios servicios
Específicos: referentes de un solo servicio

3.1. Según estén o no especificados: implícitos o explícitos

- **Criterios implícitos:** son aquellos en los que no se especifica cuáles son los criterios de calidad que se valoran; se perfilan en función de los conocimientos del sujeto que realiza la evaluación. No es recomendable utilizar este tipo de criterio de forma exclusiva. Pueden emplearse más como mecanismo de detección de problemas o de segunda revisión de los casos clasificados de calidad deficiente. Ejemplo de criterio implícito: evaluar si el diagnóstico es correcto sin que se especifique cuál es el concepto de buen diagnóstico que emplea el evaluador.
- **Criterios explícitos:** son aquellos en los que se describe en qué consisten. Se aconseja utilizar este tipo de criterio. Permite tener claro cuáles han sido los referentes para la medición, que pueden ser criticados o justificados, si es preciso. Ejemplo de criterio explícito: la toma de tensión arterial se realizará con esfigmomanómetro de mercurio, con el paciente en reposo y sin que haya fumado ni tomado café en 30 minutos previos a la medición.

3.2. Según su origen: empíricos o normativos

- **Criterios empíricos:** no dicen cómo tienen que ser las cosas para ser de calidad. Describen cómo es la realidad y se acepta como válido lo más frecuente. Por ejemplo, dar como referente de la estancia media adecuada en el hospital para una determinada patología la media de los hospitales de una determinada área o región.
- **Criterios normativos:** dicen explícitamente como tienen que ser las cosas, según la evidencia científica existente, para que sean de calidad aceptable. Son generalmente los más adecuados.

3.3. Según su dependencia o no de otros criterios: secuenciales o transversales

- **Longitudinales, secuenciales o encadenados:** el criterio secuencial es aquel que

precisa para su evaluación de los datos o cumplimiento de otro criterio anterior. Se evalúa después del precedente. Su distinción es importante por las implicaciones que tienen para su medición; más concretamente para los esquemas de muestreo, al tener que asegurarnos un número mínimo de casos para poder evaluar todos los criterios, tal como veremos en el capítulo 5. Por ejemplo: para evaluar el *Screening* de cáncer de cérvix podemos tener lo siguiente: 1er criterio: toda mujer entre 35 y 65 años deberá tener realizada la citología y registrado el resultado (normal o no) en la historia clínica. 2º criterio: si la citología es anormal, la mujer deberá ser derivada al ginecólogo para el estudio de seguimiento. Para que se pueda evaluar el 2º criterio previamente se ha de evaluar y cumplir el 1º; el 2º criterio es un criterio secuencial o encadenado al 1º.

- **Transversales:** se evalúan todos de forma independiente el uno del otro.

3.4. Según que se ponderen o no: isovalentes o ponderados

- **Isovalentes:** a todos los criterios se les da el mismo valor como requisitos de calidad.
- **Ponderados:** al resumir la evaluación se le da a algunos criterios más valor que a otros. Ejemplo: criterio 1: a todo paciente hipertenso se le realizará un ECG en el año del diagnóstico. Criterio 2: a todo paciente hipertenso se le realizará anamnesis sobre antecedentes familiares de hipertensión, cardiopatía, y dislipemia. Criterio 3: Todo paciente hipertenso se considera bien controlado si su TA \leq 140/90 en los tres últimos controles. Si nosotros a la hora de evaluar decimos que el criterio 3 vale el doble que el 1 y el 2, estamos ante un criterio ponderado; si por el contrario, todos nos parecen igual de importantes, serían criterios isovalentes. La forma más frecuente de ponderación es, al aplicar un listado cerrado de criterios, hacer que todos sumen 100 o la unidad y ponderar cada uno en función del porcentaje o proporción que se supone importa sobre la valoración global que se haga.

La ponderación suele ser arbitraria y discutible, por lo que en general es preferible evitarla.

3.5. Según el número de servicios a los que incluyen

- **Genéricos:** incluyen a varios productos o servicios diferentes.
- **Específicos:** solamente afectan a un solo servicio o producto.

Esta distinción es semejante a la que vimos a la hora de hablar de problemas de calidad en el capítulo 1. De hecho, si el problema es genérico, va a tener que ser evaluado con toda probabilidad con criterios genéricos. Algunos ejemplos: “no deben existir reacciones adversas a medicamentos” (o con un cierto estándar), sería criterio genérico de calidad de prescripción; si lo referimos a un tipo de enfermedad concreta sería específico: “no debe haber reacciones alérgicas a la penicilina en el tratamiento de la amigdalitis”.

4. ¿Cómo se construye un criterio?

El proceso de elaboración y selección de criterios es algo que atañe a todos los implicados en su eventual cumplimiento. Todo el personal sanitario debe participar en la selección de criterios, en la medida en la que estén involucrados en el servicio a evaluar, y en particular en el cumplimiento de los criterios que se decida utilizar en la evaluación.

En instituciones grandes no es operativo que todo el personal participe en la selección de criterios y ésta puede delegarse al grupo de mejora. Para ello pueden utilizarse técnicas de toma de decisiones en grupo. Se ahorra tiempo si se elabora un borrador previo sobre los criterios que hay que discutir y luego se consensuan con el resto de los profesionales.

A continuación se sugieren algunas recomendaciones para la elaboración de criterios:

- Manejar criterios simples.
- Limitar el número de criterios (el menor número posible. Máximo 10 criterios).
- Incluir sólo los mínimos parámetros esenciales.

Los criterios pueden clasificarse en función a diversas características. De todas ellas destacamos como más adecuados los criterios explícitos, normativos, isovalentes y específicos.

La elaboración o adaptación de criterios es conveniente que sea participativa y asegurándose su idoneidad por el equipo que lo va a utilizar, así como su validez y fiabilidad.

- Elaborar criterios que se adapten a los recursos de los profesionales y el entorno.
- Asegurar el consenso y la aceptabilidad por parte de quienes los tienen que cumplir.
- Asegurarse de que el contenido esté actualizado y sea válido.
- Prestar atención cuidadosa a las excepciones y aclaraciones de los términos empleados (explicados más adelante en este capítulo).
- Asegurar la fiabilidad mediante pilotaje, sobre todo si va a haber más de un evaluador.

Previo a la elaboración de criterios hay que revisar y actualizar la bibliografía existente, consensuar con el grupo de trabajo, ver si son factibles y realizables en nuestro medio y con nuestras características peculiares. Copiar criterios ya elaborados sin tener en cuenta estas reflexiones puede resultar contraproducente para el ciclo de mejora en el que se quieren emplear.

Si hacemos referencia a las clasificaciones de criterios que vimos en el apartado 3, en la elaboración de criterios, debemos tender a que sean explícitos, normativos, isovalentes y específicos, que son los que resultan en general más útiles y entendibles para iniciar actividades de mejora de la calidad.

5. Excepciones y aclaraciones a los criterios

Una primera y elemental precaución para que los criterios sean válidos y fiables (es decir, que sean interpretados de igual manera por todos los evaluadores) es definir lo más exhaustivamente posible las excepciones y aclaraciones que los acompañan.

5.1. Excepciones a los criterios

Hay circunstancias en las cuales no es exigible el cumplimiento del criterio, a estas circunstancias se las llama *excepciones al criterio*. Por ejemplo, si el criterio es “los pacientes con otitis media aguda deberán ser tratados con amoxicilina y clavulánico en 1ª instancia”, una clara excepción sería pacientes alérgicos a la penicilina. Está claro que aunque la amoxicilina es el fármaco de elección, no debe darse a un paciente alérgico. Otro ejemplo frecuente

son las excepciones a los criterios de diagnóstico, en caso de que el diagnóstico sea antiguo o haya sido realizado en otro nivel.

Las excepciones al criterio, no deben ser muchas ni frecuentes y deben de tratarse con cuidado a la hora de evaluar la calidad, ya que pueden ser un factor de confusión para las comparaciones; sobre todo si las consideramos como cumplimiento del criterio, como veremos más adelante.

5.2. Frecuencia de las excepciones

Si a la hora de evaluar un criterio vemos que aparecen muchas excepciones, debemos eliminar ese criterio; ya que corremos el riesgo de que no tenga validez como criterio de calidad o que dificulte la evaluación exigiendo diseños muestrales complicados. Por ejemplo; un criterio puede ser: debe constar en la HC de toda gestante que se da consejo sobre tabaco en el embarazo. Excepción: paciente en cuya historia consta que no fuma. A la hora de evaluar o medir el número de veces en que se da consejo a una gestante sobre tabaco, si son muchos los casos en que aparecen excepciones, no sabremos realmente si se da el consejo en un número suficiente de casos (porque tenemos una muestra menor, solamente de aquellas que realmente fuman). Probablemente habría que rediseñar el criterio dirigiéndolo solamente a las que fuman o encadenándolo a la exploración sobre el hábito tabáquico.

Si tenemos muchas excepciones, corremos el riesgo de terminar no midiendo aquello para lo que diseñamos el criterio.

5.3. Las excepciones ¿son cumplimientos o incumplimientos de los criterios?

Otro problema que plantean las excepciones es la forma de valorarlas en la evaluación: Si a la hora de la evaluación, las excepciones que aparecen las consideramos como cumplimiento del criterio, estamos perdiendo la información *real* de lo que se está haciendo (no estamos midiendo aquello que queríamos medir). Volviendo a nuestro ejemplo del consejo antitabáquico en la embarazada, si evaluamos una HC de una paciente gestante que no fuma (excepción), y lo consideramos como cumplimiento

del criterio, la evaluamos igual que si fuera una gestante fumadora a la que se dio consejo antitabáquico. En el primer caso no estamos valorando el cumplimiento del criterio al dar consejo antitabáquico a la gestante (no estamos midiendo aquello que queríamos medir). Por otra parte, si no consideramos a las excepciones como cumplimientos y les damos el mismo tratamiento que si no cumple el criterio, tampoco refleja la realidad de los hechos, ni deja ninguna opción de mejora. En nuestro ejemplo, si en esas mujeres que no fuman decimos que no cumplen con el criterio, tampoco estamos reflejando fielmente la realidad porque seguramente no se les dio consejo por no ser parte del grupo de riesgo.

¿Qué decisión adoptar? Lo que parece más lógico, en general, es considerarlas como cumplimiento de los criterios; sin embargo, lo más recomendable es considerarlas aparte. De no hacerlo así, pueden ser un importante *factor de confusión* para las comparaciones entre evaluaciones y entre centros; piénsese que un cumplimiento alto de un criterio podría ser debido a una proporción alta de excepciones, más que a la realización concreta del criterio en sí; podría incluso darse la paradoja de que en los pocos casos reales del criterio a medir, no se hiciera correctamente, mientras que en otro centro o servicio con menos excepciones el cumplimiento real fuese mayor, pero aparecer disminuido por la baja proporción de excepciones. Un ejemplo claro, frecuente y real es comparar la calidad de los diagnósticos de padecimientos crónicos, considerando como cumplimiento de los criterios si el diagnóstico es antiguo o se ha realizado en otro nivel; la diversa proporción de estas excepciones entre los diferentes centros a comparar, confunde claramente el criterio de calidad que queremos medir, si lo consideramos como cumplimiento.

5.4. Importancia de las aclaraciones

Se entiende por aclaración al criterio la descripción inequívoca de cualquier término que aparezca en el criterio o sus excepciones que pueda estar sujeto a interpretaciones. Para garantizar el empleo exacto y repetible de los términos técnicos, hay que definirlos explícitamente. Hay que dar una definición lo sufi-

cientemente detallada para que la persona que evalúa no tenga que hacer ningún juicio personal sobre si un caso cumple o no un criterio determinado. Por ejemplo, un criterio sería: “los pacientes con urocultivo positivo deberán ser tratados en los tres días posteriores al cultivo. Excepción: “paciente tratado en la visita inicial. Aclaración: urocultivo positivo significa ± 100.000 o 105 colonias/cc registrado en el informe del cultivo”. De esta forma, el evaluador no tiene que interpretar qué significa “urocultivo positivo”

Ejemplo: criterio de exploración física: “tener realizada una auscultación cardiopulmonar y pulsos periféricos”. Al no haber aclaraciones, este criterio también puede ser problemático. Si sólo tiene realizada la auscultación y no los pulsos ¿cumple el criterio?; ¿es suficiente poner normal o un símbolo alusivo o hay que detallarla?; ¿es suficiente con una localización?

Criterio de seguimiento de pacientes con alto riesgo cardiovascular: “tener realizado e informado un ECG en los dos últimos años”. Sin aclaraciones, no está lo suficientemente claro: ¿qué se entiende por ECG *informado*?; ¿es suficiente poner normal o patológico?

La existencia o no de aclaraciones, cuando estas son necesarias, y la suficiencia de las mismas repercute directamente en la fiabilidad de los criterios. La fiabilidad podemos definirla como el grado de reproducibilidad de los resultados cuando el criterio es utilizado por observadores diferentes. Criterios sujetos a interpretación diversa por distintos evaluadores no van a poder ser utilizados de forma fiable para comparar evaluaciones ni dentro de los centros ni entre centros.

5.5. Inclusión de las excepciones y aclaraciones en el diseño de los criterios

Las excepciones pueden influir en la fiabilidad y en la validez de los criterios como factores de confusión y las aclaraciones influyen en la fiabilidad. Validez y fiabilidad son dos atributos importantísimos para los criterios a medir. En consecuencia, además de que consideremos adicional y específicamente tanto validez como fiabilidad, es muy aconsejable utilizar un formato para la descripción de criterios que inclu-

Es mejor, en general, considerar a las excepciones como casos aparte, no como cumplimientos del criterio. De lo contrario, pueden ser un importante factor de confusión al comparar mediciones.

Las aclaraciones correctas y exhaustivas evitan que la persona que evalúa tenga que “interpretar” el criterio; de este modo las aclaraciones contribuyen a que el criterio sea fiable.

La validez es el principal atributo exigible a los criterios. El análisis de su presencia en un criterio puede realizarse considerando diversos tipos de validez de entre los que vamos a destacar la validez facial, de contenido y de criterio.

La validez facial o aparente hace referencia a la relevancia del criterio a primera vista. Debe ser obvia su importancia sin demasiadas explicaciones.

ya explícitamente excepciones y aclaraciones. En caso de no ser necesarias se consignará así, pero es conveniente asegurarse de que, al menos, reflexionamos sobre la necesidad de constatarlas. Este formato de descripción de criterios que incluye excepciones y aclaraciones es el que se describe en la figura 4.2. Si se utiliza es más probable que construyamos criterios válidos y fiables, pero es conveniente considerar aparte y explícitamente estas dos características antes de dar por aceptado el criterio.

6. Validez

Un criterio válido es aquel que mide aquello que está pensado que mida; es decir, aquel que mide realmente la calidad del servicio evaluado, de forma que su presencia es equivalente a presencia de calidad, y sus diferentes valores o niveles son sinónimos también de diferentes niveles de calidad. En general, para ser válido debe estar clara su influencia en el resultado de la atención (cumplimiento de expectativas y necesidades de los usuarios del servicio) y deben de tenerse en cuenta los posibles factores de confusión (que hacen variar por sí solos el cumplimiento del criterio, sin que responda esta variación a actuaciones realmente diferentes).

La validez es probablemente el principal atributo exigible a un criterio. Para evaluar la calidad, no se trata de tener muchas medidas de validez y fiabilidad dudosa, sino de disponer de medidas clave válidas y fiables. La validez es un concepto ligado al error sistemático: si el criterio no es válido vamos a estar sistemáticamente midiendo una cosa diferente a la prevista.

6.1. Validez facial, de contenido y de criterio

La validez facial, lógica o aparente hace referencia a que el criterio parezca importante para el problema a evaluar sin que sea preciso dar muchas explicaciones; es su grado de relevancia obvia. No es un aspecto de la validez que determine la validez real, pero sí es una buena guía para priorizar y excluir criterios obviamente no válidos. Los criterios tienen que tener que ver con el problema de calidad a evaluar. Por ejemplo, el estado vacunal o los antecedentes de cáncer en la familia no parece que sean relevantes como criterios de calidad para la anamnesis en el hipertenso.

La validez de contenido hace referencia a que el criterio mida el concepto que se quiere medir; en nuestro caso, sería asegurarnos que está en relación con la calidad; es decir, que se refiere a algunas de las dimensiones con las que hemos definido calidad, y se relaciona con las necesidades y expectativas de los usuarios. La valoración de este tipo de validez no dejará de ser, sin embargo, de tipo fundamentalmente subjetivo, igual que la de la validez facial, pero también nos puede servir para descartar criterios no claramente relacionados con el concepto o aspecto de calidad que queramos medir. Por ejemplo, si no hemos incluido la satisfacción del profesional como una de las dimensiones del servicio que se ofrece al cliente externo, un criterio que la mida no tiene validez de contenido para evaluar la calidad del servicio que recibe el usuario externo. De igual manera, si queremos medir la calidad científico-técnica, no serán válidos criterios que hagan referencia

Criterio	Excepciones	Aclaraciones
Descripción del referente que se va a utilizar como medición (contenido del criterio)	Especificación de las situaciones en las que es aceptable que el criterio no se cumpla	Definición detallada de cualquier término del criterio, las excepciones, o de algo que pueda ser interpretado de más de una forma a la hora de evaluar

Figura 4.2. Formato para la descripción de criterios

a aspectos estructurales que tengan relación básicamente con la accesibilidad.

La validez de criterio es el grado en que la variable escogida se correlaciona con un criterio de referencia objetivo, fiable y que esté ampliamente aceptado como una buena medida del fenómeno de interés. En nuestro caso sería la relación probada con un mejor resultado, si es un criterio de proceso, con diferencias en el proceso si es un criterio de resultado y con factibilidad del proceso si es un criterio de estructura. En ausencia de una medida de referencia, valorar una adecuada validez es un tema complejo. Muchas veces resulta imposible usar este abordaje por falta de estudios que lo avalen. Sin embargo, es de una gran importancia y debemos procurar siempre reflexionar sobre ella. La forma de realizarlo es evaluando la evidencia científica que justifica los criterios. Lo ideal es que pueda estar especificado el método empleado para revisar las evidencias científicas cuantitativas en las que se fundamenta; y que consten las fuentes de información utilizadas. Para evaluar el nivel de evidencia científica se han desarrollado diferentes sistemas de graduación que se basan generalmente en el tipo y calidad de los estudios de los que se derivan las recomendaciones clínicas o decisiones más adecuadas en los distintos pasos del proceso que se hayan analizado. Dado que estas mismas recomendaciones, o decisiones más correctas, son las que podemos utilizar para construir los criterios que evalúen la calidad de la atención, estos mismos sistemas de valoración de la evidencia científica de las recomendaciones clínicas nos van a ser útiles para referenciar la validez de los criterios que midan la realización o no de esas recomendaciones.

Actualmente existen numerosos sistemas de valoración de la evidencia científica para respaldar la fuerza de las recomendaciones clínicas. Una revisión realizada en el marco de la AHRQ estadounidense (<http://archive.ahrq.gov/clinic/epcs/sums/strengthsum.pdf>) encontró 109 documentos que proponían algún tipo de sistema para valorar la calidad y el grado de evidencia correspondiente de los artículos científicos, aunque muchos de ellos referidos a un determinado tipo de estudio, y un total de siete sistemas de gradación general de la fuerza de la evidencia. Hay sistemas que se

proponen como útiles para la valoración de recomendaciones sobre decisiones terapéuticas, de diagnóstico, preventivas, de pronósticos, etc., pero en general la base común es valorar el tipo y rigor de los estudios en que se basan. En este sentido, uno de los primeros sistemas de valoración de la evidencia científica, aún de utilidad, es el que propuso la Canadian Task Force on Preventive Health Care (cuadro 4.3).

Esta clasificación fue modificada, ampliándola en la AHRQ, quedando como se muestra en el cuadro 4.4.

De todas formas, no es un tema cerrado y sigue trabajándose sobre ello a nivel internacional. Merece la pena, en este sentido, citar la

Cuadro 4.3 Clasificación del nivel de evidencia científica	
I.	Evidencia obtenida de al menos un ensayo diseñado apropiadamente, aleatorizado y controlado.
II.	Evidencia a partir de estudios no aleatorizados o no controlados. II.1.- Evidencia obtenida de ensayos controlados bien diseñados, sin aleatorización. II.2.- Evidencia obtenida de estudios analíticos bien diseñados de cohortes o caso-control preferiblemente de más de un centro o grupo de investigación. II.3.- Evidencia obtenida de series múltiples de tiempo con o sin la intervención. Los resultados muy llamativos en experimentos no controlados pueden ser considerados también como evidencia de este tipo.
III.	Opiniones de expertos basadas en experiencia clínica, estudios descriptivos o informes de comités de expertos. Éste es el grado más bajo de validez.

Cuadro 4.4 Clasificación del nivel de evidencia propuesto por la AHRQ	
Ia.	La evidencia científica procede a partir de meta-análisis de ensayos clínicos controlados y aleatorizados.
Ib.	La evidencia científica procede de al menos un ensayo clínico controlado y aleatorizado.
IIa.	La evidencia científica procede de al menos un estudio prospectivo controlado, bien diseñado y sin aleatorizar.
IIb.	La evidencia científica procede de al menos un estudio casi experimental, bien diseñado.
III.	La evidencia científica procede de estudios descriptivos no experimentales, bien diseñados como estudios comparativos, de correlación o de casos y controles.
IV.	La evidencia científica procede de documentos u opiniones de expertos y/o experiencias clínicas de autoridades de prestigio.

iniciativa GRADE, que parece estar convirtiéndose en un referente.*

Obviamente, es muy poco práctico analizar de forma exhaustiva la validez de todos los criterios a utilizar; sin embargo, es necesario al menos reflexionar sobre ello y cuestionarse aquellos que se hacen exclusivamente por consenso. Por otra parte, muchas de las decisiones a tomar en los procesos asistenciales no están, no ya probadas sino ni siquiera estudiadas; no es incorrecto entonces basarse en el consenso para elaborar criterios, pero sabiendo que serían en cierto modo criterios *provisionales*.

Cuando lo que vamos a explorar no son criterios ya probados como requisitos de calidad, sino lo que hemos llamado causas hipotéticas, el propio diseño del estudio nos va a informar (al medirlas concurrentemente con el criterio o indicador que representa el problema a evaluar) si tienen al menos validez interna; es decir, para el entorno en que se realiza el ciclo de mejora.

Una vez hemos reflexionado sobre la validez, hasta dónde nos sea posible, lo que hemos de asegurarnos es que los criterios a emplear sean *fiabiles*. Tenemos que considerar que el indicador o criterio que elijamos para representar el problema a evaluar es efectivamente válido y fiable, antes de la medición.

6.2. Fiabilidad

Como ya hemos indicado más arriba, la fiabilidad podemos definirla como el grado de reproducibilidad de los resultados cuando el criterio es utilizado por observadores diferentes. Criterios sujetos a interpretación diversa por distintos evaluadores no van a poder ser utilizados de forma fiable para comparar evaluaciones ni dentro de los centros ni entre centros. Una descripción sin ambigüedades de todos los términos empleados en la definición del criterio o indicador, así como una especificación clara de las excepciones y forma de medición, contribuyen a que los criterios o indicadores sean fiables. Sin embargo, es conveniente, o incluso imprescindible, que la fiabilidad sea comprobada empíricamente, antes de generalizar el uso del criterio o indicador. Particularmente si va a ser

utilizado por varios evaluadores o para más de una evaluación.

Para medir la fiabilidad se utilizan diversos índices de acuerdo, generalmente fáciles de calcular e interpretar: índice kappa, índice de concordancia general e índice de concordancia específica. Si el criterio es fiable, estos índices deben ser elevados. El más extendido es probablemente el estadístico kappa de Cohen que se usa para corregir el acuerdo debido al azar y permite estimar la significación estadística. Vamos a ver cómo se calculan e interpretan cada uno de ellos.

6.3. Índice Kappa

El índice kappa (k) mide la concordancia total que existe si excluimos la debida al azar, o el acuerdo real más allá del azar. Su cálculo responde a la fórmula general:

$$k = \frac{P_o - P_e}{1 - P_e}$$

donde P_o es la concordancia observada, y P_e es la concordancia esperada debida exclusivamente al azar. El valor de k va de 0 a 1, aunque pueden resultar también valores negativos, siendo 0 el valor que representa la concordancia por azar, y 1 el valor de concordancia máximo, que respondería a una fiabilidad perfecta.

Una vez calculados P_o y P_e , el cálculo de k puede realizarse según la fórmula indicada por Fleiss tanto para parejas de evaluadores como para el caso de evaluadores múltiples. Existen otras alternativas, al parecer estadísticamente más correctas que las fórmulas de Fleiss, para el caso de que sean más de dos evaluadores, pero considerablemente más complejas.

Veamos un ejemplo de cálculo del índice kappa para una pareja de evaluadores:

- Para obtener los datos necesarios, el criterio cuya fiabilidad queremos conocer es utilizado de forma independiente por dos evaluadores para evaluar las mismas historias clínicas. Vamos a suponer que seleccionamos una muestra de 30; con los resultados de los dos evaluadores podemos construir una tabla en la que se reflejan las veces que han coincidido en que el

La fiabilidad se mide con diversos índices como el kappa, la concordancia general y la concordancia específica.

* Dichas propuestas pueden consultarse en su página web: <http://www.gradeworkinggroup.org/>

criterio se cumple o que no se cumple. La tabla, ya con los datos obtenidos en nuestro ejemplo simulado, sería la siguiente:

Evaluador 1	Evaluador 2		
	Cumple	No cumple	
Cumple	9	1	10
No cumple	6	14	20
	15	15	30

Esto quiere decir que, según el evaluador 1, 10 historias cumplen el criterio y 20 no; mientras que para el evaluador 2, 15 historias cumplen el criterio y 15 no. Las celdas de la tabla indican la distribución de los que cumplen y no cumplen según cada evaluador, en función de lo que dijo el otro. Así por ejemplo, de los 20 casos en los que el evaluador 1 dijo que el criterio no se cumplía, en 14 coincidió con el evaluador 2, y en 6 no coincidió porque el evaluador 2 dijo que sí cumplía el criterio.

La concordancia observada (P_o) es el total de coincidencias, dividido por el total de casos evaluados. Según los datos de la tabla

$P_o = \frac{9 + 14}{30} = 0.77$ es decir un 77%. ¿Y cuál sería P_e , la concordancia esperada por puro azar? P_e se calcula sumando el producto de las proporciones de cumplimiento que han dicho los dos evaluadores al producto de las proporciones de incumplimiento.

En nuestro ejemplo, según el evaluador 1, la proporción de cumplimiento es

$$\frac{10}{30} = 0.33$$

Según el evaluador 2 la proporción de cumplimiento es

$$\frac{15}{30} = 0.5$$

El producto de ambas = $0.33 \times 0.5 = 0.165$

Según el evaluador 1, la proporción de incumplimiento es

$$\frac{20}{30} = 0.67$$

Según el evaluador 2, la proporción de incumplimiento es $\frac{15}{30} = 0.5$

El producto de ambas = $0.67 \times 0.5 = 0.335$ es igual a la suma de ambos productos = $0.165 + 0.335 = 0$

Con ello podemos calcular el índice Kappa

$$\frac{P_o - P_e}{1 - P_e} = \frac{0.77 - 0.5}{1 - 0.5} = 0.54$$

¿Qué podemos concluir ante valores como este?, ¿es el criterio fiable o no? En esencia, ¿qué valores de kappa son adecuados? Como respuesta, existen diversas escalas de valoración en la literatura pero en general, la mayoría de los autores coinciden en dar por aceptables kappas mayores de 0.4. Para ser buenos deberían de ser mayores de 0.6. Por ejemplo, según criterio de Fleiss se considera que un kappa traduce buena fiabilidad si es mayor o igual a 0.6. Según Landis y Kock, un kappa negativo o igual a cero indica acuerdo bajo o inexistente, hasta 0.2 acuerdo ligero, por encima de 0.4 acuerdo moderado, a partir de 0.6 acuerdo notable, a partir de 0.8 acuerdo alto y 1 acuerdo absoluto. Según todos los autores, los kappas por debajo de 0.4 no son aceptables. El kappa de nuestro ejemplo sería aceptable.

6.4. Circunstancias a valorar en la interpretación del índice Kappa

A pesar de su extendido uso, el índice kappa no está exento de inconvenientes. El valor y la interpretación del índice kappa deben expresarse con algunas cautelas en relación con diversos factores que pueden hacerlo variar con independencia de la fiabilidad real del instrumento (criterio o indicador) valorado. Estos factores son la prevalencia de la cualidad que el instrumento mide, la sensibilidad y especificidad de los diversos evaluadores, y la propia adecuación de la fórmula aplicada cuando se analiza la fiabilidad de más de dos evaluadores simultáneamente.

La influencia de la prevalencia (o proporción verdadera de presencia de la característica evaluada) en el valor de kappa es que para prevalencias (expresadas en nuestro caso como

Las prevalencias extremas hacen que el índice kappa tienda a ser cero y su valor no se puede interpretar en la forma habitual.

Para interpretar adecuadamente el índice kappa, hay que tener en cuenta la prevalencia de lo que mide el criterio que se evalúa.

El índice kappa es relativamente sencillo de calcular, con base en el conocimiento de la concordancia observada y la esperada por azar. Para ser aceptable ha de tener un valor mayor de 0.4; para ser bueno ha de ser mayor de 0.6

Unas correctas especificaciones de los criterios (excepciones, aclaraciones, mecánica de evaluación) ayudan a evitar los problemas de las distintas sensibilidad y especificidad de los evaluadores.

proporción de cumplimientos de los criterios) cercanas a 1 o a 0, kappa tiende a ser cero con independencia del grado de acuerdo entre los evaluadores, a menos que éste sea total, en cuyo caso el valor de kappa para prevalencia de 1 o 0 es infinito. A la vista de los cálculos simulados realizados por diversos autores para diversas prevalencias, manteniendo constante en cada caso la sensibilidad y especificidad de los evaluadores, parece evidente que, como regla práctica general, el valor de kappa no debe ser interpretado de la forma habitual para prevalencias menores de 0.15 o mayores de 0.85

En nuestro caso, como puede ser el caso en la práctica la mayoría de las veces en que se pilote y evalúe la fiabilidad de los criterios o indicadores para medir la calidad, la prevalencia es inicialmente desconocida; lo que proponemos hacer, es estimarla con base en calcular la media de las prevalencias (proporciones de cumplimiento) que propone cada evaluador. Está claro que lo ideal es conocer previamente la prevalencia (proporción de cumplimientos real); sin embargo, aunque puede haber considerables desviaciones en relación al valor real, las situaciones extremas van a ser identificadas. En resumen y como norma, una correcta interpretación de un índice kappa bajo necesita del conocimiento, o al menos de un cálculo estimado, de la prevalencia de lo que valora el instrumento analizado; el valor de kappa sólo deberá ser concluyente por sí solo si es alto: en kappas bajos debemos asegurarnos de que no estamos ante prevalencias extremas.*

La sensibilidad y especificidad de los evaluadores en relación a lo evaluado, generalmente no es posible conocerla ni estimarla de antemano, pero es, sin embargo, un factor también de potencial influencia en los valores de kappa. De hecho, según cálculos simulados publicados, si la sensibilidad de un evaluador (capacidad de ver la existencia de la cualidad o criterio si realmente está) fuese inferior al 70%, el valor de kappa va a resultar necesariamente inferior a 0.4, porque la baja sensibilidad conduce por sí misma a discordancias. La manera más común de normalizar (y elevar) la capacidad discriminatoria de los evaluadores es entrenándolos. En ausencia de entrenamiento, la

mejor medida es construir el instrumento de valoración con unas especificaciones para su manejo (excepciones, aclaraciones, mecánica de evaluación) lo suficientemente detalladas para que sea necesariamente fiable en las condiciones normales de aplicación.

El cálculo del valor de *kappa para múltiples evaluadores* merece también un comentario. El hecho de que la valoración del cumplimiento de los criterios la efectúen más de dos observadores aumenta la complejidad del cálculo y se pueden utilizar modelos jerárquicos, o bien la fórmula de Fleiss para más de dos observadores. La fórmula propuesta por Fleiss es cuestionada por algunos autores, pero la complejidad del cálculo de las fórmulas alternativas para varios evaluadores, junto al hecho de que el kappa para dos evaluadores sigue sin cuestionar, aconseja basar nuestros cálculos rutinarios en los valores de kappa por parejas. Si el pilotaje se realiza con más de dos evaluadores, puede calcularse kappa para las diversas combinaciones de parejas.

Finalmente, la correcta interpretación del índice kappa precisa del cálculo de su intervalo de confianza, de forma que daremos por aceptables sólo los kappas cuyo límite inferior del correspondiente intervalo de confianza sea mayor de 0.4. El lector interesado puede encontrar la fórmula para calcular el intervalo de confianza para el índice kappa en el libro de Fleiss, incluido en la bibliografía de este capítulo; también puede dirigirse al sitio web.*

Después de todas estas aclaraciones y consideraciones sobre las particularidades del índice Kappa vamos a ver los otros índices para valorar la fiabilidad.

6.5. Índice de concordancia general

Indica la concordancia de los evaluadores al clasificar el criterio. Se define como la proporción de casos en que el juicio de ambos evaluadores coincide. Se calcula dividiendo el número de acuerdos de ambos observadores por el total de casos evaluados. Es equivalente a Po o proporción de acuerdo observado, en la fórmula del cálculo de kappa.

Para su interpretación aislada para valorar la fiabilidad hacen falta valores muy altos para que

* Una explicación detallada, así como la fórmula de ajuste del kappa por la prevalencia estimada puede verse en: <http://www-users.york.ac.uk/~mb55/msc/clinimet/week4/kappa.htm#seci> (consultado en noviembre de 2015).

sea considerada buena. Según proponen Palmer y Banks concordancias $\geq 95\%$ van a ser siempre sinónimos de herramientas fiables. Estos niveles de concordancia son exigibles, así como kappas como mínimo buenos (>0.6) cuando el criterio o indicador se va a generalizar y utilizar para comparar diversos centros en programas externos.

6.6. Índice de acuerdo específico

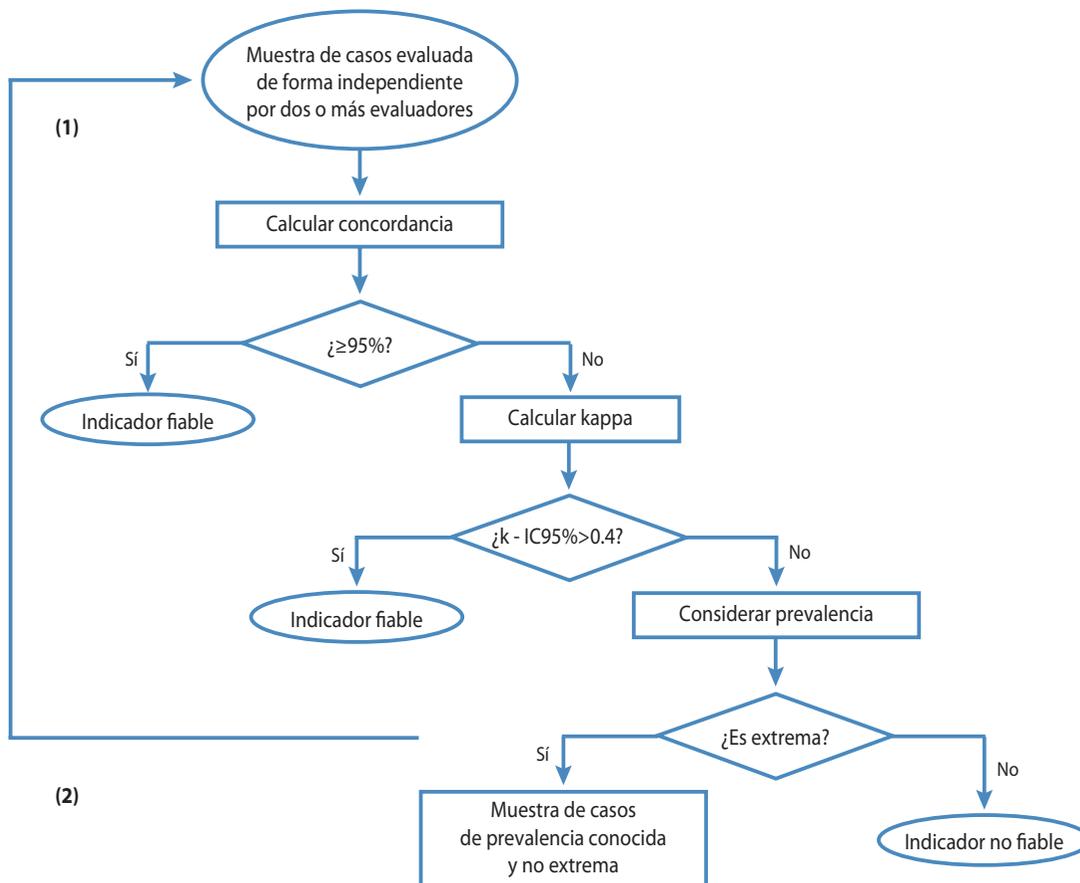
Es el acuerdo al clasificar como negativo un criterio (concordancia al decir que no se cumple). Se calcula dividiendo el número de casos en que ha habido acuerdo al decir que el criterio no se cumple, por el número de casos a los que al menos un observador ha dicho que no se cumple. Tiene interés cuando nos interesa especialmente que no se clasifique como incumplidores a quienes puedan estar haciéndolo bien. Su interpretación es semejante a la de la concordancia general.

6.7. Procedimiento práctico para evaluar la fiabilidad

Una forma simple, intuitiva y pragmática para la valoración de la fiabilidad es basarla inicialmente en el cálculo de la concordancia simple, y sólo si ésta no es alta ($<95\%$) calcular el índice kappa, ya que en ocasiones concordancias sólo moderadamente altas van unidas a kappa aceptables. Además, el índice kappa debería acompañarse, como hemos visto, de un conocimiento previo o una estimación de la prevalencia de la cualidad que evalúa el instrumento cuya fiabilidad queremos analizar.

La figura 4.3 contiene un esquema, en forma de algoritmo, que proponemos para valorar la fiabilidad de los criterios e indicadores, con base en la discusión que acabamos de exponer.

El índice de concordancia general es la concordancia observada. Son exigibles valores muy elevados (mayores o iguales a 95%) para utilizarlo con seguridad como medida aislada de fiabilidad.



La valoración de la fiabilidad puede exigir la consideración de la concordancia, el kappa y la prevalencia, en una secuencia práctica que explicita la figura 4.3.

(1): punto de inicio si la prevalencia es desconocida
 (2): punto de inicio si conocemos previamente la prevalencia. Alternativamente puede ajustarse el valor kappa por la prevalencia estimada
 IC: intervalo de confianza

Figura 4.3. Procedimiento de análisis de la fiabilidad de los indicadores para medir calidad

BIBLIOGRAFÍA

- Fleiss JL. Statistical Methods for rates and proportions. New York: John Willey & Sons, 1981. Cap 13.
- Hernández Aguado I, Porta Serra M, Miralles M, García Benavides F, Bolúmar F. La cuantificación de la variabilidad en las observaciones clínicas. Med Clin (Barc) 1990; 95: 424-429.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.
- Palmer RH. Evaluación de la asistencia ambulatoria. Principios y práctica. Madrid: Ministerio de Sanidad y Consumo, 1990. Cap. 8.
- Saturno PJ. Cómo diseñar una evaluación sobre la calidad asistencial en atención primaria. En: IV Congreso de la Sociedad Andaluza de Medicina Familiar y Comunitaria. Sevilla: Consejería de
- Salud, 1992.
- Thompson WD, Walter SD. A reappraisal of the kappa coefficient. J Clin Epidemiol 1988; 41(10): 949-958.

Anexo 4.1

Trabajo práctico recomendado: la construcción de criterios para evaluar la calidad. Análisis de validez

Se recomienda realizar un análisis de al menos un criterio para razonar y justificar su validez, siguiendo las siguientes indicaciones:

Paso 1. Construcción de criterios

Para comenzar, construir el listado de criterios que definen la calidad del servicio o aspecto del servicio a evaluar utilizando el formato que incluye excepciones y aclaraciones (figura 4.2 de este capítulo). A continuación, seleccionar un criterio del listado que se propone. Consignar este criterio seleccionado en el formato que se adjunta, en el cual se irán anotando también los resultados de los pasos sucesivos del trabajo práctico que son los siguientes.

Paso 2. Reflexionar sobre la validez facial, de contenido y de criterio

- Validez facial: ¿se entiende la relevancia del criterio a primera vista? Justificar brevemente por qué son importantes los criterios elegidos para la calidad del servicio o aspecto del servicio a evaluar.
- Validez de contenido: ¿se refiere claramente a algunas de las dimensiones del concepto de calidad?, ¿su cumplimentación satisface algunas de las necesidades o expectativas de los usuarios?
- Validez de criterio: ¿conocemos alguna fuente o estudio en donde se demuestre la evidencia científica del criterio? ¿qué tipo de evidencia es, según la clasificación señalada en este capítulo?

Paso 3. Excepciones y aclaraciones.

- ¿Son suficientes? En caso negativo, completar el criterio elegido hasta que contenga las excepciones y aclaraciones que se consideren necesarias para que tenga el máximo de objetividad.
- Si se va a realizar el trabajo en grupo, discutir y acordar con todos los implicados hasta lograr una descripción definitiva de los criterios seleccionados, con excepciones y aclaraciones.
- Pensar en situaciones en las que es aceptable que el criterio no se cumpla (excepciones) y si todos los términos del criterio y las excepciones se pueden interpretar de forma unívoca a la hora de emplearlas para evaluar.

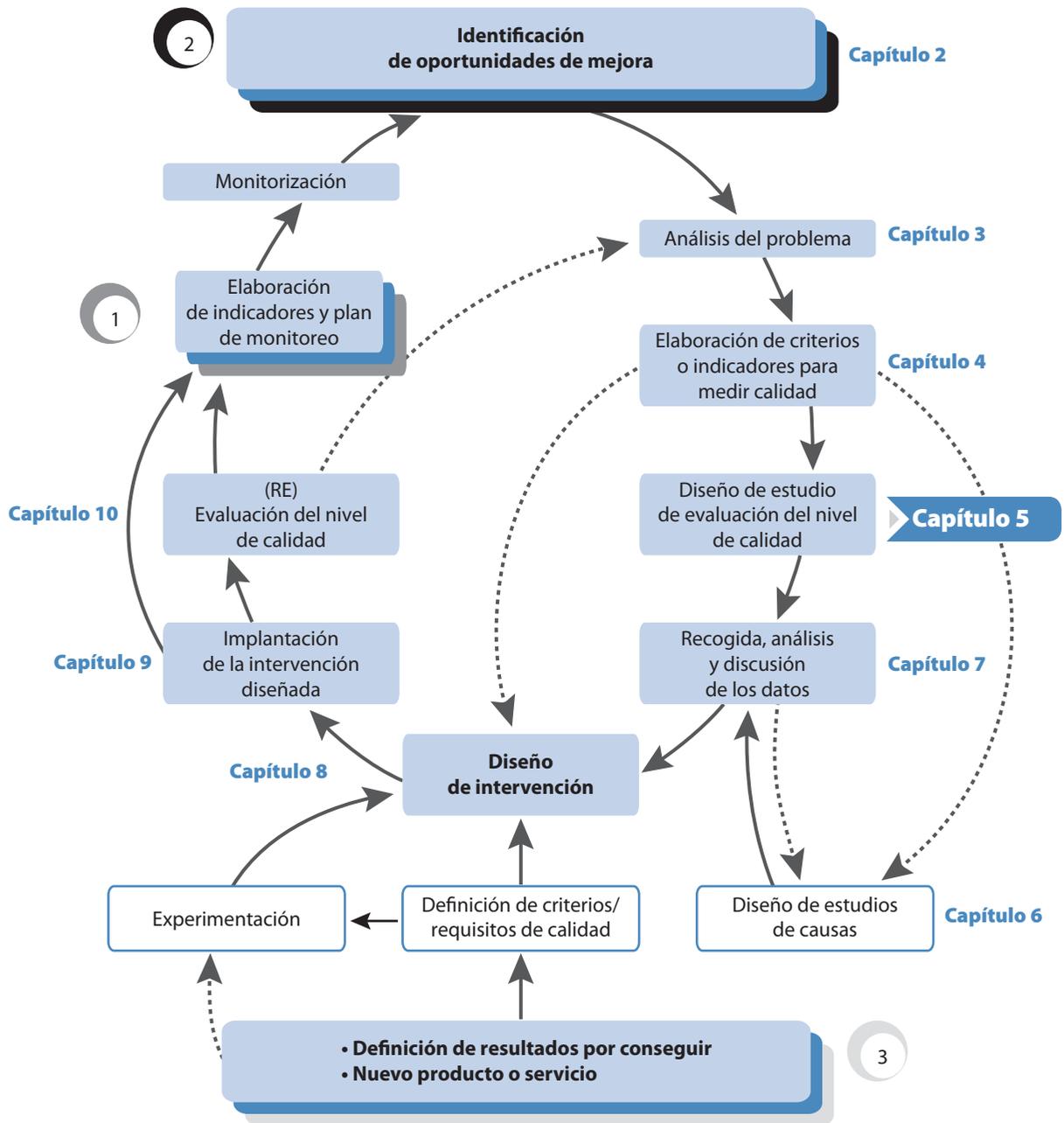
Capítulo 5

Diseño de estudios de nivel de calidad y la investigación de causas. Componentes de una evaluación

Contenido general y objetivos específicos

1. Introducción
 2. Estudios del nivel de calidad y estudios de causas de la calidad deficiente
 3. Diseño de estudios del nivel de calidad: componentes por especificar
 4. Criterios, dimensiones y tipos de datos
 5. Definición de las unidades de estudio
 6. Fuentes de datos
 7. Identificación y muestreo de las unidades de estudio. Determinación del marco muestral. Ventajas e inconvenientes del muestreo
 8. Número de casos necesarios. Representatividad, precisión y confianza de la muestra
 9. Métodos de muestreo
 10. Muestreo aleatorio simple. Metodología, ventajas e inconvenientes
 11. Muestreo sistemático. Metodología, ventajas e inconvenientes
 12. Muestreo estratificado. Metodología, ventajas e inconvenientes
 13. Muestreo de conveniencia. Metodología, ventajas e inconvenientes
 14. Tipo de evaluación, según relación temporal con la acción evaluada
 15. Tipo de evaluación, según quién toma la iniciativa
 16. Tipo de evaluación, según quién obtiene los datos
- Anexo 5.1. Número de casos necesarios para asegurarse un determinado intervalo de confianza de 95%, según diversos valores esperados de la variable que se mide
- Anexo 5.2. Intervalo (i) para un grado de confianza de 95%, según estimación puntual y tamaño muestral
- Anexo 5.3. Práctica recomendada. Formato para el diseño del estudio de nivel de calidad

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Puntos de partida



Vía de acción normal



Vía contingente o alternativa



1 Monitorización



2 Ciclo de mejora



3 Diseño

Contenido general

Una vez establecidos los criterios o requisitos de calidad de los aspectos por evaluar, hemos de diseñar cuidadosamente su medición. Veremos en este capítulo que los estudios de evaluación de la calidad pueden tener diversos niveles de complejidad, pero que podemos agruparlos en dos que denominamos, a falta de una terminología mejor, estudios del nivel de calidad y estudios de causas de calidad deficiente. Se explica más en detalle el primero de ellos, que consiste en medir el nivel de presencia de los criterios o requisitos de calidad preestablecidos, característico de las evaluaciones de la calidad de la atención clínica. Se ofrece asimismo un esquema detallado de los componentes del estudio que deben quedar bien definidos sea cual sea el servicio o producto objeto de evaluación. En otro capítulo se abordan las características de los estudios de causas, más adecuados para problemas de tipo organizativo o gerencial, en los que no hay más criterios, requisitos o indicadores de calidad que el que represente el problema que se va a analizar.

Objetivos específicos

1. Diferenciar los diversos tipos de estudio que pueden diseñarse para evaluar la calidad, en especial, aquéllos que miden requisitos de calidad, por un lado, y aquéllos que buscan causas de calidad deficiente, por el otro.
2. Definir las características de las unidades de estudio (casos) en las que se va a realizar la medición.
3. Distinguir entre marco temporal para el muestreo de los casos y periodo de tiempo en el que los criterios deben cumplirse.
4. Determinar si todos los criterios se miden en idénticas unidades de estudio o no.
5. Definir cuáles van a ser las fuentes de los datos que se necesitan en el estudio.
6. Determinar el número de casos que se van a evaluar.
7. Valorar las ventajas e inconvenientes de utilizar diversos tamaños de muestra.
8. Explicar en qué consiste una muestra representativa.
9. Distinguir muestreos aleatorios de muestreos de conveniencia.
10. Escoger el método de muestreo que más se ajuste al contexto de su centro.
11. Desarrollar paso a paso la metodología para extraer la muestra.
12. Definir quién debe encargarse de recoger los datos.
13. Resumir, en el diseño del estudio, las decisiones tomadas en relación con la dimensión de la calidad que se va a evaluar y los tipos de datos que se van a utilizar.

Diseño de estudios de nivel de calidad y la investigación de causas. Componentes de una evaluación

1. Introducción

En previos capítulos, hemos tomado decisiones importantes y se han resumido las implicaciones teóricas y prácticas en relación con los primeros tres pasos de los ciclos de mejora: la identificación del problema de calidad u oportunidad de mejora (capítulo 2), el análisis del problema seleccionado (capítulo 3) y la elaboración de criterios válidos y fiables para medir la calidad (capítulo 4). Este capítulo aborda el paso siguiente: diseñar adecuadamente la recogida de datos para estudiar y cuantificar el problema de calidad.

2. Estudios del nivel de calidad y estudios de causas de la calidad deficiente

Hay dos grandes grupos de estudios cuyos resultados nos van a servir para decidir qué hacer para mejorar: los estudios del nivel de calidad y los estudios de causas hipotéticas de la calidad deficiente. Con los primeros medimos el grado de cumplimiento de los criterios o requisitos de calidad previamente definidos, cuyo grado de presencia o ausencia sabemos que equivale al grado (o nivel) de calidad de lo que vamos a evaluar. Normalmente, éste es el caso de las evaluaciones de calidad de *problemas de atención clínica*, con los cuales comprobamos si estamos haciendo las cosas de forma correcta. En estos estudios medimos directamente lo que sabemos que son requisitos (transformados en criterios o indicadores) de calidad.

Una situación diferente ocurre cuando el problema no es de tipo clínico sino *organizacio-*

nal o gerencial (tiempo de espera, descoordinaciones en el equipo, procesos innecesariamente complicados, etc.). Podemos incluso tenerlo ya medido en términos de lo que representa el problema de calidad, pero para su mejora tenemos que averiguar las causas que lo provocan. Esta situación también puede darse después de los estudios de nivel de calidad, si no sabemos qué hacer para mejorar un determinado criterio clínico que no se cumple. En ambos casos, lo que tenemos es una serie de causas hipotéticas que investigar, pero no la certeza de que su presencia o ausencia vaya unida a la presencia o ausencia de calidad, como ocurre en los estudios que hemos llamado del nivel de calidad. Por ejemplo, si determinamos que el tiempo de espera para la consulta es criterio de calidad y queremos medir la proporción de pacientes que esperan más de un cierto tiempo, diseñaríamos un estudio de nivel de calidad. Si, por el contrario, ya sabemos que el tiempo de espera es excesivo y no tenemos idea de por qué es así, diseñaríamos un estudio para averiguar, de todas las causas hipotéticas (dependientes del paciente, de la organización del centro, de los horarios, de las normas, etc.), cuáles se asocian a tiempos de espera prolongados en nuestro caso concreto. En los estudios de causas hipotéticas es necesario medir concurrentemente con las posibles causas, el criterio o los criterios que nos representan la calidad del aspecto en estudio. Por ejemplo, no podemos averiguar las causas de los tiempos de espera prolongados si no medimos a la vez las causas por indagar y los tiempos de espera a los que se asocian. Por esto, el estudio de causas debe ir precedido de la definición de los

En este capítulo se avanza un paso más en el ciclo de mejora: el diseño de estudio para medir calidad.

En los estudios de causas hipotéticas no sabemos si éstas influyen o no en la calidad.

En los estudios del nivel de calidad medimos criterios de los cuales se sabe de antemano que están relacionados con la calidad.

En este capítulo nos ocuparemos del diseño de un estudio de nivel de calidad.

criterios o indicadores de calidad del problema en estudio; estos mismos criterios o indicadores son los que tendremos que medir tras la intervención para asegurarnos de que hemos mejorado.

En este capítulo, se detallan los componentes de un estudio de nivel de calidad con el objetivo de averiguar hasta qué punto se cumplen o no los criterios de calidad elegidos; sabremos de antemano que la ausencia de estos criterios es directamente causa de calidad deficiente. Los aspectos metodológicos son semejantes en los estudios de causas hipotéticas en lo que se refiere al estudio del nivel de calidad del que se parte, aunque difiere el análisis de datos para el cual necesitamos ver, en el caso de los estudios de causas hipotéticas, la relación del nivel de calidad con las diversas causas hipotéticas en estudio. Los estudios de causas hipotéticas son el objeto del capítulo 6.

3. Diseño de estudios del nivel de calidad: componentes por especificar

Los estudios sobre calidad han de ser una actividad rigurosa, planificada y factible. Para asegurarnos de que es así, resulta útil seguir una cierta sistemática, que nos ayude a considerar todos los aspectos que son relevantes en su realización. Adicionalmente, la especificación homogénea de los componentes del estudio sirve para poder replicarlo y para comparar evaluaciones posteriores de un mismo centro o realizadas en centros distintos.

El esquema que proponemos tiene sus raíces en el Programa Ibérico, modificado con base en experiencias posteriores. Consta de una serie de componentes, la mayoría de los cuales están interrelacionados; sin embargo, conviene considerarlos aisladamente y procurar su definición exhaustiva. Una vez definidos todos, queda también definido el diseño de la evaluación. Estos componentes son:

1. Criterios para evaluar la calidad
2. Dimensión estudiada
3. Tipos de datos
4. Unidades de estudio
5. Fuentes de datos
6. Identificación y muestreo de las unidades de estudio
 - marco muestral

- número de casos
 - método de muestreo
7. Tipo de evaluación
 - según relación temporal
 - según quién tomó la iniciativa
 - según quién obtiene los datos

Varios de estos componentes han sido vistos en capítulos anteriores y haremos una breve reseña de los mismos; revisaremos los demás con mayor detalle.

4. Criterios, dimensiones y tipos de datos

Para diseñar nuestra evaluación, hay que empezar por definir estos tres componentes. Los criterios, traducción de los requisitos de calidad, han sido tratados en el capítulo 3 y las dimensiones, en el capítulo 1. Es aconsejable revisar estos capítulos, aunque a continuación vayamos a recordar los puntos más relevantes para el diseño del estudio.

4.1. Criterios

Los criterios son la herramienta con la cual mediremos la calidad. Su aplicación nos sirve para discernir entre buena y mala calidad y entre los diferentes niveles de ésta. Hay que recordar que, en la práctica, aceptar unos criterios determinados, equivale a establecer y a aceptar asimismo unos determinados objetivos de calidad en la asistencia que prestamos. Un criterio es a la vez un requisito de calidad que queremos que se cumpla; de ahí la tremenda importancia de que sean relevantes, realistas y con posibilidad de ser aceptados sin controversia por todos los profesionales implicados. Una correcta elaboración y selección de criterios es, junto a una priorización adecuada del tema que se va a evaluar, el principal factor determinante del éxito o del fracaso del ciclo de mejora.

En el capítulo 4 revisamos una serie de características que hay que considerar en relación con los criterios. En este punto es conveniente revisarlas de nuevo y explicitar las que definitivamente queremos medir, como el primero de los componentes del diseño de nuestro estudio. Es importante que no olvidemos considerar su validez, las excepciones y las aclaraciones pertinentes.

Se propone un esquema de componentes que conviene definir separada y exhaustivamente.

El primer componente es la descripción de los criterios que se van a utilizar, especificando las excepciones y haciendo las aclaraciones necesarias.

4.2. Dimensiones

A la vista del aspecto que se va a evaluar y de los criterios que vamos a medir, hemos de describir cuál de las varias dimensiones de la calidad vamos a valorar.

Como vimos en el capítulo 1, proponemos el uso de un listado de seis dimensiones (efectividad, eficiencia, seguridad, atención centrada en el paciente, accesibilidad equidad) o de uno reducido de cuatro dimensiones (calidad científico-técnica, seguridad, accesibilidad y atención centrada en el paciente).

La mayoría de las evaluaciones que realizan los profesionales incluyen la medición, muchas veces como única dimensión, de algún aspecto de la calidad científico-técnica. Esto es así no sólo porque es la dimensión más amplia y con más subcomponentes; también influye el que sea la que se ve más cercana a la actuación de los profesionales y sobre la que más directa y fácilmente podemos actuar. Es también la dimensión más comúnmente aceptada como indiscutible para caracterizar la calidad de la atención prestada. Sin embargo, no debe olvidarse la importancia de garantizar asimismo unos niveles adecuados de accesibilidad y satisfacción: accesibilidad tanto interna (en el sentido de la existencia o inexistencia de barreras a la continuidad de cuidados) como externa, si pensamos que actuamos sobre poblaciones y no exclusivamente sobre pacientes que nos visitan; y atención centrada en el paciente unida a la satisfacción, como un resultado por conseguir para garantizar la adecuada y necesaria colaboración de los pacientes y población en los programas y tratamientos.

En el diseño del estudio hay que hacer explícita la dimensión que hemos decidido medir. Con ello aclaramos también cuál es la dimensión concreta que intentamos mejorar y avisamos implícitamente sobre las dimensiones que van a escapar a nuestra medición; es decir, sobre las que no podremos afirmar si están o no presentes, ni si conseguiremos o no mejorar. Naturalmente, se asume que, aunque no haya cambios en las otras dimensiones, la mejora de una de ellas supone un salto cualitativo en el servicio que ofrecemos.

4.3. Tipos de datos

Los criterios por utilizar van a estar referidos a datos de estructura, proceso o resultado. Hacerlo explícito como uno de los componentes del diseño permite recordar la necesaria reflexión sobre los problemas de validez que puedan plantear cada uno de ellos. También es conveniente constatarlo para dejar claro que sabemos de la existencia de otros datos potencialmente relevantes que no se tomarán en cuenta.

La distinción de estructura, proceso y resultado, propuesta por Donabedian y profusamente utilizada en el sector salud, no forma parte de la terminología utilizada normalmente en sectores de la industria pero reviste un gran interés por las diferentes precauciones que hay que tomar en cuanto a su validez, y para tener argumentos en las discusiones que periódicamente aparecen dando preferencia a uno u otro tipo de datos. El resumen de estos tres tipos de datos se describe en el cuadro 5.1.

El segundo componente es aclarar qué dimensión o dimensiones (calidad científico-técnica, accesibilidad, atención centrada en el paciente, seguridad) habremos de evaluar.

El tercer componente es la constatación explícita del tipo de datos (estructura, proceso, resultados) que vamos a utilizar.

Cuadro 5.1

Validez, facilidad de obtención y utilidad inmediata de los diferentes tipos de datos

Tipo de dato	Validez	Facilidad de obtención	Utilidad inmediata
Estructura	Es condición necesaria pero no suficiente	+++	+++
Proceso	Depende, ver relación con el resultado	++	+++
Resultado	- Depende, ver relación con el proceso - Es preciso ajustar para eliminar factores de confusión	+	+

Las unidades de estudio se delimitan definiendo las características de la población que recibe la atención evaluada, los proveedores de la atención y el tiempo en que la atención tiene que haber sido prestada a cada paciente.

5. Definición de las unidades de estudio

La descripción de la unidad de estudio para la evaluación es el equivalente a la definición de caso por evaluar. Se trata, en esencia, de determinar cuáles son las características de la unidad básica, más desagregada, sobre la que vamos a efectuar la medición de los criterios de calidad asistencial.

Habitualmente, tendremos que especificar tres aspectos:

- *Características del paciente o población que entra en la evaluación*, coincidente con el grupo población objeto o receptor de la atención que evaluamos. Este grupo puede ser definido en función de su edad, sexo, pertenencia a determinado grupo de riesgo, diagnóstico de determinada enfermedad, sometido a determinado tratamiento o intervención, etc. Por ejemplo, en el caso de la evaluación de la atención preventiva del cáncer femenino, podemos decidir que será objeto de estudio la atención recibida en este sentido por toda mujer mayor de 18 años residente en nuestra área de cobertura, o bien que haya acudido al menos una vez a nuestro centro sanitario; para la detección de hipertensos, todos los adscritos a nuestro centro y que pertenezcan a ciertos grupos de edad, etcétera.
- *Características de los profesionales y nivel asistencial evaluado*. Por ejemplo, un médico de un determinado centro o servicio, todos los médicos de un centro o servicio, el personal de enfermería, el personal de fisioterapia, todo el equipo de atención primaria, todos los centros de un área, etcétera. Como en el caso anterior, esta característica es excluyente: sólo vamos a considerar aquellos casos poblacionales de características determinadas que además hayan dependido, en relación con el problema, de una determinada instancia o nivel asistencial, que es el que evaluamos.
- *Periodo de tiempo evaluado*. Coordinada que no hace referencia necesariamente al tiempo físico convencional (un año determinado, un mes, etc.), sino al segmento

del proceso asistencial, tiempo de contacto o relación entre cada caso que entra en el estudio y el nivel asistencial evaluado, en el cual queremos que se vea cumplido el criterio de calidad que vamos a medir. Por ejemplo, podemos definir que en nuestro programa de detección precoz de cáncer de mama, la exploración de mama debe realizarse en la primera visita de toda mujer incluida el grupo de edad que hemos definido. De este modo, el periodo de tiempo que define nuestra unidad de estudio es una visita, y más concretamente, la *primera visita* a nuestro centro, de manera que al obtener los datos para nuestra evaluación no vamos a fijarnos en las actuaciones realizadas en visitas sucesivas. En otros casos, especialmente en el seguimiento de enfermedades crónicas, podemos plantearnos evaluar la atención prestada en un año, en las visitas siguientes al diagnóstico, etc. Otros criterios pueden no tener referencia temporal, como es el caso de muchos de los criterios de estructura.

En una evaluación con varios criterios, es perfectamente posible que cada uno de ellos se refiera a periodos de tiempo o segmentos del proceso diferentes.

Normalmente se define un marco temporal más amplio (por ejemplo, el último año de funcionamiento del centro), para acotar de manera general la extracción de casos para la evaluación pero que *no define* el periodo evaluado en cuanto a la atención prestada en el problema concreto que queremos evaluar. Hay que hacer un esfuerzo por reflexionar ante cada criterio si lleva o no un periodo de tiempo concreto, dentro del proceso asistencial, en el que evaluamos si se cumple o no. Puede coincidir o no con el marco temporal que definimos para el marco muestral, es decir, para la extracción de casos por evaluar.

La experiencia nos ha mostrado que este aspecto de la definición de la unidad de estudio es el más difícil de entender y especificar, especialmente cuando se quiere evaluar la atención en casos de pacientes crónicos.

Para evitar la confusión entre el periodo de tiempo como marco temporal para definir el universo de donde extraer los casos (un año, seis meses, etc.) y el periodo de tiempo como fragmento del proceso asistencial en el que vamos a comprobar si el criterio se cumple (una visita, un cierto tiempo tras el diagnóstico, etc.) es conveniente reflexionar sobre ello criterio a criterio.

6. Fuentes de datos

En nuestro diseño tenemos que hacer explícito de dónde va a ser posible extraer los datos que necesitamos para la evaluación. En evaluaciones retrospectivas, muy probablemente vamos a utilizar las historias clínicas o cualquier otro sistema o soporte de registro de datos en relación con el tema evaluado. Otra fuente de datos frecuente (a veces obligada por el tema o por la inexistencia de registros sobre el dato de interés, o porque lo que nos interesa es lo que puede decirnos el usuario) son las encuestas *ad hoc*. En otras ocasiones (habilidades manuales, técnicas de entrevista clínica, etc.), la evaluación sólo es posible mediante observación directa.

En cualquier caso, una regla que hay que seguir siempre es procurar comenzar diseñando criterios que puedan ser cualificados con los datos o sistemas de registro ya existentes. En aras de la factibilidad (y su correspondiente repercusión en el éxito de la implantación de las actividades de GC), hay que procurar no empezar con evaluaciones que precisen el montaje de nuevos registros o la realización de encuestas complicadas.

Desafortunadamente, el relativo déficit de buenas fuentes de datos limita con frecuencia las posibilidades y la extensión de las evaluaciones de la calidad asistencial. En relación con ello, es común encontrar, como la primera o una de las primeras actividades de GC en los centros sanitarios, la evaluación de la calidad de los registros e historias clínicas. Sin quitarle mérito a esa actividad, pensamos que, una vez garantizada la existencia de registros, es más efectivo que los déficits que tengan aparezcan evidentes en función de su repercusión en la buena o mala asistencia que se presta a un problema de salud determinado (por ejemplo, registro de edad y sexo para saber si la persona es susceptible de entrar en un Programa de detección precoz de cáncer femenino, registro de peso y talla y resultados de la analítica en pacientes cuyo diagnóstico, tratamiento o seguimiento adecuados dependen precisamente del conocimiento explícito de esos valores, etc.); de hecho los registros han de ser contempla-

dos como un medio más para lograr una buena asistencia y no como un fin en sí mismos.

7. Identificación y muestreo de las unidades de estudio. Determinación del marco muestral. Ventajas e inconvenientes del muestreo

Una vez definida la unidad de estudio, nos hemos de plantear los mecanismos para identificar los casos de pacientes o población que responden a la definición de caso, y el marco muestral o universo de donde extraeremos, en su caso, una muestra representativa de éstos para su evaluación. Los mecanismos de identificación de casos y la facilidad relativa para realizar un muestreo varían según el tipo de atención por evaluar y según las características de cada centro.

Frecuentemente, sobre todo en los casos en los que la atención que se va a evaluar se encuadra en un determinado programa o grupo bien definido (hipertensión, diabetes, encamados en domicilio, niño sano, pacientes con cervicalgia etc.), existen ficheros específicos que facilitan el muestreo e identificación de los casos por evaluar. Otras veces, es posible utilizar a tal efecto los registros de morbilidad, o cualquier otro registro rutinario en el que conste explícitamente la característica que define la unidad de estudio. Otras veces, el universo para muestreo e identificación de casos han de ser directamente las historias clínicas o un muestreo aleatorio de la población cubierta o atendida. Todo depende del tema concreto que se va a evaluar, la definición que hagamos de la unidad de estudio y las características estructurales de cada centro.

Es aconsejable tener presentes, muy en primer término, las posibilidades de identificación de casos y obtención de los datos pertinentes al escoger el tema y diseñar los criterios para evaluar: en primera instancia es preferible abordar evaluaciones con criterios aplicables a casos de identificación posible y fácil, aunque resulten menos atractivas que otras para las que el proceso de identificación de casos pueda resultar costoso o complicado.

A continuación hay que describir de qué fuente vamos a identificar los casos (listados, libros de registro, censos, etc.) y de dónde extraer los datos para la evaluación. Hay que evitar, en lo posible, crear fuentes de datos.

Es aconsejable tener muy claro en el diseño la delimitación del universo o marco muestral y la factibilidad de los mecanismos para la obtención de los casos a evaluar y los datos para la evaluación.

Evaluar con muestreo es una manera eficiente y útil de tener datos, que se convierte en obligatoria si el universo de casos o marco muestral es muy grande.

En las evaluaciones de calidad solemos utilizar inicialmente muestras manejables, relativamente pequeñas (50 a 60 o hasta 100 casos) pero no exentas de rigor científico.

El tamaño de muestra decidido para evaluar niveles organizativos superiores (por ejemplo, centro sanitario) no sirve para evaluar niveles más desagregados (por ejemplo, cada médico del centro).

Si el total de casos (universo o marco muestral) que responde a la definición de unidad de estudio es pequeño o disponemos de toda la información en base de datos informatizados, podemos plantearnos la posibilidad de evaluarlos todos; sin embargo, por lo general es más conveniente extraer una muestra.

El muestreo, aparte de ser el proceder de elección cuando el universo de unidades de estudio es relativamente grande, y obligatorio cuando es muy grande, tiene la ventaja de que es menos costoso (por lo menos en tiempo) que evaluar la totalidad de casos, y permite obtener resultados útiles a corto plazo. Los inconvenientes, si es que puede llamárselos así, son, por un lado, que para hacerlo correctamente hay que manejar con cierta soltura unos conceptos estadísticos básicos; por otro, que puede haber escépticos que no terminen de dar crédito a los resultados por el simple hecho de ser una muestra. Es muy aconsejable prepararse para afrontar estos dos inconvenientes porque las ventajas superan con mucho los esfuerzos que puede suponer el poder evaluar con muestras.

8. Número de casos necesarios. Representatividad, precisión y confianza de la muestra

La definición del número de casos o unidades de estudio necesarios para la evaluación es uno de los puntos que más angustia produce al diseñar una muestra. Es también un punto en el cual, el enfoque aconsejable para los estudios de evaluación de la calidad puede diferir sustancialmente del recomendado para trabajos de investigación de otro tipo. Podemos adelantar que un tamaño de muestra manejable y habitualmente suficiente es de 50 a 60 casos, y como máximo 100, pero conviene entender por qué y aprender a extraer la muestra de forma que sea representativa. Para ello, tras decidir cuál es el nivel o unidad organizativa para la cual queremos extraer conclusiones, hay que tener claros los conceptos de confianza, precisión y representatividad de la muestra.

8.1. Unidad organizativa para la que queremos extraer conclusiones

Al definir la unidad de estudio en lo relativo al proveedor del servicio que se evalúa, hemos

decidido si nos interesa que sea un profesional individual, un equipo, todo el centro, etcétera; con ello hemos contribuido a definir también el marco muestral. Sin embargo, esta decisión tiene además una gran trascendencia para determinar el tamaño de la muestra: el número de casos que decidamos extraer nos va a permitir obtener conclusiones sobre la calidad al nivel organizativo que previamente hayamos decidido, pero no a un nivel inferior. Es decir: una muestra pensada para evaluar la calidad de atención al nivel de centro sanitario no va a servir para evaluar a cada profesional por separado. Si queremos evaluar cómo lo hace cada médico, enfermero o fisioterapeuta, vamos a necesitar el mismo tamaño de muestra pensado para el centro, pero para cada uno de los profesionales individuales. Es factible la agrupación de muestras pensadas para niveles organizativos más bajos para evaluar niveles superiores (por ejemplo, agrupar los datos de cada médico para evaluar el centro), pero no al revés. Hay que tener esto muy claro y muy en cuenta al diseñar el estudio.

8.2. Representatividad, precisión y confianza de la muestra

Estos tres conceptos, con estos y otros nombres, suelen considerarse relacionados con el tamaño de la muestra y se utilizan a veces de forma confusa e intercambiable, aunque son diferentes. Vamos a aclarar su significado.

Supongamos que estamos interesados en saber la altura media de los pacientes que están en un determinado día en las salas de espera de un centro de salud grande. Si quisiéramos saberlo con toda exactitud, mediríamos la talla de todos ellos y calcularíamos la media; en este caso, como no es una muestra, no habría problema de representatividad, precisión ni confianza del resultado. Pero, ¿y si son cientos de personas? Podemos considerar, entonces, la posibilidad de hacer una *estimación* midiendo la talla sólo en una muestra de estas personas. Al ser una muestra, el resultado que obtengamos está sujeto a errores, pero esos errores pueden conocerse y controlarse de manera que los tengamos en cuenta a la hora de dar los resultados de nuestro muestreo. Estos errores son de tres tipos y se relacionan con la representatividad, la precisión y

la confianza de los resultados o estimaciones de la muestra.

El primer error que podemos tener es hacer una selección *sesgada* de las personas que incluimos en la muestra. Por ejemplo, supongamos que seleccionamos todos de las consultas de pediatría. Obviamente, esta muestra no es *representativa* de los pacientes que esperan en las consultas del centro de salud, y esto *con independencia* de cuántos niños hayamos incluido en la muestra. El problema, el error, es que en nuestra muestra los niños están sobrerrepresentados, hayamos medido muchos o pocos. Para ser *representativa* tendríamos que haber hecho la selección de casos para la muestra de manera que todos los pacientes de todas las consultas hubieran tenido la misma oportunidad de entrar en la muestra, y no únicamente los de pediatría. El control del error de sesgo para que la muestra sea representativa depende del método de muestreo, no del tamaño de la muestra. La muestra va a ser *representativa* (no sesgada) si se ha realizado de forma *aleatoria*, lo cual equivale a decir que se le da a todas las unidades del universo muestral la misma oportunidad de entrar en la muestra. Otra cosa es, una vez que nos aseguramos de que la muestra no es sesgada, en qué grado las estimaciones que hagamos con nuestra muestra responden o no al valor real que tiene, en el universo muestral, lo que estamos midiendo con una muestra. Es decir, siguiendo con nuestro ejemplo, hasta qué punto podemos inferir la altura media real de todos los que están esperando en las consultas a partir de una muestra representativa (no sesgada) de los mismos. Para saberlo, tenemos que tener en cuenta los otros dos errores que, éstos sí, varían con el tamaño de la muestra.

Una de las cosas que nos va a preocupar es la *precisión* de la estimación. Sabemos que el valor que encontramos en la muestra puede no ser el real pero, ¿y si, a partir de la estimación de la talla media que hemos obtenido en la muestra y teniendo en cuenta la variación (dispersión) de valores diferentes y el tamaño de la muestra, tuviéramos un medio para afirmar que el valor real se encuentra entre unos determinados límites? Así podríamos conocer la magnitud de este error. Pues bien, el medio de calcularlo existe, y la forma de reconocerlo y controlarlo es dar el resultado de la mues-

tra con un *intervalo* de valores, dentro del cual sabemos (por la teoría de la probabilidad) que está el valor real de lo que estamos midiendo, aunque no sepamos a ciencia cierta cuál de los valores incluidos en el intervalo es el real (¡para eso tendríamos que medir a todos, no a una muestra!). Cuanto más pequeño sea el intervalo que nos resulta, más *precisa* será nuestra estimación, es decir, más pequeño será este error de muestreo. Lógicamente, la estimación va a ser más precisa con intervalo de posibles valores reales más pequeño (menos error) cuanto más grande sea la muestra. En nuestro ejemplo, a cuanto más gente midamos la talla, más se acercará nuestra estimación al valor real; los cálculos matemáticos oportunos nos mostrarían que podemos dar nuestra estimación con intervalos de valores posibles cada vez más pequeños.

Pero aún nos queda otra posibilidad de error: ¿hasta qué punto estamos seguros de que el valor real está realmente comprendido en el intervalo de valores que hemos calculado? Está claro que sólo tendremos 100% de certeza si medimos a todos los pacientes que esperan. Como hemos obtenido una muestra, tenemos una cierta probabilidad de equivocarnos si afirmamos que los resultados obtenidos en la muestra serían los que obtendríamos si midiéramos a todos los pacientes. Esta probabilidad de estar o no equivocados al afirmar que el valor real es uno de los valores del intervalo que hemos calculado para saber la precisión, es lo que se conoce como significación o *grado de confianza* de la estimación que hemos obtenido con el muestreo. La confianza en el resultado del muestreo aumentará, naturalmente, si aumentamos el tamaño de la muestra, pero también si decidimos dar el resultado con menos precisión (con un intervalo de valores más amplio). Pensemos en nuestro ejemplo: si decimos que, según los resultados de la muestra representativa (es decir, sin sesgo, extraída de forma aleatoria), la altura media estimada es de 160 cm y que la media real está en un intervalo de ± 20 cm en torno a esa media estimada (es decir, entre 140 y 180 cm), tendremos menos posibilidades de equivocarnos (más confianza) para un mismo tamaño de muestra, que si afirmamos que el valor real está entre 160 cm ± 5 cm (es decir, entre 155 y 165).

La representatividad, la precisión y la confianza representan tres tipos de errores diferentes que podemos tener al medir con muestreo.

La representatividad hace referencia al error de sesgo y no depende del tamaño de la muestra sino del método de obtención de la misma.

La precisión de los datos medidos en una muestra se refiere a un error de muestreo que podemos calcular y reconocer dando el resultado de la evaluación con un intervalo de valores, dentro del cual está el valor real.

La precisión aumenta (es decir, el intervalo de valores posibles disminuye) a medida que aumenta el tamaño de la muestra.

La confianza es una medida de otro posible error del muestreo y representa la probabilidad de que el resultado que hemos obtenido con la muestra (precisión incluida) sea cierto. Esta confianza o probabilidad de certeza se expresa en porcentaje y, como valor aceptable, se suele establecer 95%.

Dado que ambos errores (representados por la precisión y la confianza), aunque diferentes, están indisolublemente unidos y ambos dependen del tamaño de la muestra, cuando se dan los resultados de mediciones con base en una muestra se suelen (en realidad se *deben*) acompañar de la medida de estos dos errores, que es lo que se conoce conjuntamente como *intervalo de confianza* del resultado obtenido. El *intervalo* hace referencia a la precisión y se expresa como \pm una cierta cantidad, que hay que sumar y restar al valor estimado que hemos obtenido en la muestra, y la confianza hace referencia al grado de certeza del resultado que hemos obtenido (intervalo incluido) y se suele expresar en porcentaje. Este porcentaje expresa la probabilidad de que el valor real sea de *verdad* uno de los que está incluido en el intervalo; por ejemplo un intervalo de confianza de 95% quiere decir que estamos un 95% seguros de que el valor real está dentro del intervalo o, lo que es lo mismo, que la probabilidad de que el valor real sea un valor de fuera del intervalo es sólo de 5%. En nuestro ejemplo, si el resultado de la medición de la talla en la muestra representativa es de 160 cm \pm 10 cm, para una confianza de 95% (o lo que es lo mismo, el intervalo de confianza de 95% es 160 \pm 10 cm), quiere decir que hay 95% de probabilidades de que sea cierto que la talla media real está entre 150 cm y 170 cm (o 5% de probabilidades de que el valor real, el que resultaría de medir a todos, sea inferior a 150 cm o superior a 170 cm).

Una vez entendido el significado de los tres errores que hay que tener en cuenta al medir con muestra y cómo uno de ellos (el sesgo, traducido como representatividad) no depende del número de casos de la muestra sino del método de muestreo, podemos volver a la cuestión ¿cuántos casos hacen falta para que la evaluación sea útil, es decir, suficientemente precisa y confiable?

El proceder habitual en los trabajos de investigación es calcular, mediante fórmulas, el tamaño de la muestra habiendo fijado antes la precisión y la confianza, de forma que los resultados tengan el nivel de interpretabilidad buscado. Las fórmulas que se emplean con más frecuencia son las siguientes:*

$$n = \frac{z^2 \cdot p \cdot (1 - p)}{i^2}$$

Si la variable que se va a medir es cualitativa; es decir si en cada unidad de estudio se mide la presencia o ausencia de algo (una cualidad determinada), y el resultado de toda la muestra se resume en proporción o porcentaje de individuos que tienen esa cualidad; por ejemplo, cumplimiento o incumplimiento de un determinado criterio.

$$n = \frac{z^2 \cdot \sigma^2}{i^2}$$

Si la variable es cuantitativa; es decir si lo que medimos en cada individuo de la muestra es la magnitud o cantidad en que la variable en estudio está presente (por ejemplo, peso, talla, tiempo de espera, etc.), y el resultado de toda la muestra se resume como cantidad media (peso medio, talla media, tiempo de espera medio, etc.).

En estas fórmulas:

- n = tamaño de la muestra que necesitamos
- z = un número que depende de la confianza que queremos tener en el resultado. Este número representa valores de la distribución normal que se busca en las tablas correspondientes, y es 1.96 para una confianza de 95%, que es la más habitual en estudios muestrales.
- i = el intervalo o precisión, expresado en proporción (tanto por uno), para las variables cualitativas, y en unidades de medida de la variable en estudio, para las cuantitativas. Representa la precisión deseada. Por ejemplo 0.1 si queremos que el valor real esté en el intervalo comprendido entre lo que resulte en la muestra ± 0.1 (equivale a $\pm 10\%$).
- p = la proporción esperada de lo que se va a medir. Como habitualmente no se sabe, se pone 0.5 (equivalente a 50%), que es el valor de p que da un número mayor de n , para garantizar que en cualquier caso la muestra tendrá la precisión y la confianza prefijadas.
- σ = la desviación estándar que se espera va a tener la variable que se va a medir, expresada en unidades de medida de esa variable.

En los trabajos de investigación suelen emplearse fórmulas para calcular el número de casos necesarios que garantice una precisión y confianza previamente definidas.

* El lector que no esté interesado puede obviar la lectura de este apartado, puesto que no es estrictamente necesario para llevar a la práctica el contenido de este capítulo.

Si el universo o marco muestral (que se representa como N) es suficientemente pequeño (como regla, si la n que hemos calculado es mayor a 10% de N), se calcula un tamaño de muestra más reducido (n'), pero que mantiene la misma precisión y confianza. Para esto, se usa la fórmula

$$n' = \frac{n}{1 + \frac{n-1}{N}}$$

Dado que los criterios para evaluar el nivel de calidad son habitualmente variables cualitativas, se presenta un cuadro (anexo 5.1) con el número de casos necesarios para asegurarnos una confianza de 95% y diferentes intervalos (precisión), para diversas proporciones de cumplimiento esperadas (p). En este cuadro se ha destacado la columna que corresponde a la proporción esperada de 50% y el intervalos de ± 0.05 .

Sin embargo, para las evaluaciones de calidad se aconseja hacer caso omiso de estos cálculos y utilizar un tamaño de muestra manejable, que garantice la factibilidad del estudio. Aconsejamos que este tamaño inicialmente sea de 50 a 60 casos aunque pueden efectuarse evaluaciones con tamaños aún más pequeños (mínimo 30).

Utilizar esta regla práctica para determinar el tamaño de la muestra pareciera restar cientifismo a las evaluaciones de la calidad asistencial. Sin embargo, una vez realizado el muestreo aleatorio de este número relativamente pequeño de casos, nada impide enmendar la *falta de cientifismo* y calcular a posteriori el intervalo de confianza que corresponde a los valores encontrados (como veremos en el capítulo 7). Muy probablemente, estaremos suficientemente satisfechos con la información obtenida por las siguientes razones:

- *Necesidad relativa de estimaciones muy precisas*: en GC por lo general no es de primordial interés tener una gran precisión sobre el grado de cumplimiento de los criterios, sino tener una idea aproximada que ilustre sobre todo si nos encontramos muy por debajo o notablemente por encima. Intervalos de hasta ± 0.15 , es decir, ± 15 en porcentaje (o a veces más) no nos van a modificar las conclusiones sobre el tema en

relación con las necesidades de introducir medidas correctoras. Piénsese, por ejemplo, en que quisiéramos un cumplimiento de determinado criterio en al menos 75% de los casos, y que en la muestra nos sale $30\% \pm 20$; igualmente tendríamos que actuar si la proporción real de cumplimiento fuese de 10% que si fuese de 50%, que son los dos extremos del intervalo de valores posibles. El anexo 5.2 muestra los intervalos de confianza de 95% para diversos tamaños de muestra y estimaciones puntuales del nivel de cumplimiento.

- *Importancia de la factibilidad del proyecto*: en GC, tanto al inicio como para el mantenimiento de las actividades, es de gran trascendencia mantener niveles lo más altos posible de factibilidad de los proyectos. No hay que sacrificar la factibilidad a la precisión del muestreo, aunque sí combinar ambas características sabiendo que la primera va a ser determinante del éxito de la implantación rutinaria de las actividades. Con muestras relativamente pequeñas, las estimaciones pueden resultar poco precisas pero, si el muestreo está bien hecho y se ha respetado la aleatoriedad, también sabemos que la estimación nos aporta información válida.
- *Importancia de evitar muestreos ineficientes*: en las fórmulas habituales, para calcular el número necesario de casos en la muestra –por ejemplo, para una variable cualitativa que es lo más frecuente en GC–, hay que introducir no sólo la precisión que buscamos y el nivel de confianza que queremos obtener, sino también el valor estimado a priori que tiene la variable en estudio en la población de donde se extrae la muestra. Habitualmente no tenemos esta estimación al estudiar por primera vez un tema concreto en GC. Como consecuencia, es habitual calcular la fórmula *situándonos en el peor de los casos*, es decir aquel valor esperado (p) de la variable en estudio que precisaría de un número más elevado de casos para ser medida con una muestra, con resultados que tengan intervalos de confianza no más grandes que el previamente fijado. En variables cualitativas, el peor de los casos es que la variable tenga un valor estimado de

La muestra de 50 a 60 casos se extrae de forma aleatoria para que sea representativa y después calculamos el intervalo de confianza del resultado obtenido.

En las evaluaciones de la calidad, estimaciones no muy precisas dan información suficiente para decidir.

Los proyectos de mejora de calidad deben de ser factibles. Muestras grandes pueden impedir que el proyecto se lleve a cabo por el esfuerzo inicial que supone.

Utilizar en principio muestras pequeñas evita muestreos ineficientes. Respetando la aleatoriedad, siempre tendremos la opción de ampliar el tamaño de la muestra si es necesario.

Si queremos detectar diferencias pequeñas en la reevaluación, conviene utilizar muestras más grandes.

$p=0.5$ (50%). En el campo de la GC, la aplicación de esta regla, sumamente extendida en los trabajos de investigación que no pueden permitirse realizar muestreos que produzcan resultados de una precisión relativamente baja, conduce a la realización de un trabajo extra, al muestrear un número de casos mayor que el necesario para decidir, y que no se traduce en unos mejores resultados para el programa.

Si se tienen los recursos suficientes (ganas y tiempo), nada impide emplear las fórmulas habituales para calcular el tamaño de la muestra; con ello nos garantizamos una precisión determinada. Sin embargo, para una evaluación de calidad asistencial que inicia un ciclo de mejora, es preferible realizar de entrada una muestra pequeña (máximo de 100 casos), calcular a posteriori el intervalo de confianza y, en caso de que la información obtenida no resulte satisfactoria para los objetivos de GC, ampliar el número de casos de la muestra. Para esto último, ya dispondremos como mínimo de una estimación del valor de p , que va a ser mejor, en cuanto al tamaño de la muestra necesario, que *situarnos en el peor de los casos*.

8.3. Inconvenientes de las muestras pequeñas

Aparte de los problemas de precisión ya comentados, el principal inconveniente que puede tener el evaluar la calidad con base en muestras pequeñas es que, al comparar los resultados iniciales con evaluaciones posteriores, sólo vamos a encontrar significativas las diferencias importantes. Las variaciones pequeñas no van a ser detectadas o, si aparecen en los resultados, probablemente carecerán de significación estadística. Con base en lo alejados que se encuentren de nuestro estándar los primeros resultados, podremos decidir ampliar la muestra para la segunda evaluación o para ambas. Revisamos este punto en el capítulo 9.

Otra circunstancia frecuente e importante que puede complicar el muestreo es el hecho de que los diferentes criterios que vayamos a utilizar en la evaluación no sean aplicables a la misma unidad de estudio sino a *subunidades*. Por ejemplo, en el caso de utilizar criterios secuenciales, encadenados, en los cuales el primero

de ellos ha de cumplirse en todo un grupo de población (por ejemplo, detección de cáncer en mujeres de edad determinada), mientras que los siguientes aplican sólo a sectores incluidos en el primer grupo (por ejemplo, criterios de seguimiento en caso de encontrar citología o exploración mamaria positivas). En estos casos el universo de referencia para cada criterio es diferente: en nuestro ejemplo, una muestra realizada para evaluar la posible realización de las actividades de *screening*, difícilmente va a servirnos también para evaluar adecuadamente el cumplimiento de los criterios de seguimiento. Como consecuencia, puede ser conveniente realizar muestreos diferentes para el estudio de cada criterio (con referencia a su universo específico de aplicación) o realizar un muestreo amplio hasta cubrir un mínimo de 30 casos en los que pueda evaluarse el criterio aplicable al subgrupo de población más pequeño. Siempre pensando en la factibilidad de los proyectos, es aconsejable evitar en las primeras evaluaciones criterios que sólo sean aplicables a subgrupos de pacientes o población de número reducido y que no puedan ser fácilmente identificados para su muestreo.

9. Métodos de muestreo

Ya hemos señalado que la representatividad de la muestra depende del método de muestreo, y más concretamente de que éste no sea sesgado y esté hecho de manera que todos los elementos del universo o marco muestral tengan la misma oportunidad de ser incluidos en la muestra; es decir que el muestreo sea aleatorio.

Existen diversos métodos de muestreo que respetan la aleatoriedad. Los más comunes y útiles para los estudios de calidad son el muestreo aleatorio simple, el sistemático, el estratificado y combinaciones de ellos (por ejemplo estratificado y sistemático).

Los métodos que respetan la aleatoriedad nos aseguran la representatividad estadística, pero hay otros métodos de muestreo, encuadrables en lo que se conoce como muestreos de conveniencia, relativamente frecuentes y útiles pero sobre cuyas limitaciones debemos estar avisados.

Vamos a revisar las características y mecánica de todos ellos. En el diseño del estudio

En las evaluaciones con criterios secuenciales (encadenadas), es importante considerar un muestreo que nos garantice un mínimo de 30 casos para el criterio que se evalúa en el último eslabón de la cadena.

Para que la muestra sea representativa, el muestreo debe ser aleatorio.

hay que dejar bien claro cuál es el método que vamos a utilizar, a menos que midamos todos los casos, lo cual debe ser también consignado.

10. Muestreo aleatorio simple. Metodología, ventajas e inconvenientes

En el muestreo aleatorio simple, cada caso de la muestra se extrae al azar y de forma independiente. Para ello se utilizan tablas de números aleatorios u otra forma de generarlos (calculadoras, listas de lotería, etc.). Para realizarlo es preciso que todos los elementos del universo o marco muestral estén numerados y que cada número sea fácilmente localizable. La forma en que funciona es la siguiente:

1. Se determina el tamaño de la muestra (n).
2. Se dispone de un universo o marco muestral (N) previamente definido y en el que todos sus elementos estén numerados (de 1 a N) y localizables.
3. Con una tabla de números aleatorios, u otra manera de generar números aleatorios, se van seleccionando n números que han de estar comprendidos entre 1 y N .
4. Estos números corresponden a los casos que hay que extraer del universo de casos y que van a componer la muestra. Por ejemplo, supongamos que tenemos un universo de 500 historias clínicas y queremos sacar una muestra aleatoria simple de 50 historias. Para ello, generamos 50 números aleatorios comprendidos entre 1 y 500 (o entre los 500 números que le hayamos asignado a las historias) y extraeremos para evaluar las 50 historias que tengan asignadas los números aleatorios obtenidos.

El principal inconveniente del muestreo aleatorio simple es la necesidad de que el universo esté numerado y de que cada número sea fácilmente localizable (piénsese que los números aleatorios pueden saltar de una parte a otra del universo).

11. Muestreo sistemático. Metodología, ventajas e inconvenientes

El muestreo sistemático se realiza extrayendo un caso para la muestra cada cierto intervalo

fijo de casos. Este intervalo, conocido como intervalo de muestreo (k), se calcula dividiendo el número total de casos del marco muestral (N) entre el tamaño de la muestra (n). Para que sea aleatorio, *el primer caso tiene que extraerse de manera aleatoria*, de entre todo el universo o, lo que es más cómodo, de entre los primeros k casos.

En el ejemplo anterior $N=500$; $n=50$:

$$k = \frac{500}{50} = 10 \rightarrow \text{elegimos un número}$$

al azar del 1 al 10 y ésta será la primera historia para la muestra; a continuación escogemos *sistemáticamente* una historia de cada 10. Es decir, contando nueve historias entre caso y caso para nuestra muestra.

Este proceder es mucho más cómodo que el muestreo aleatorio simple, pero, dado que los casos se extraen de forma cíclica, tenemos que estar seguros de que el universo no está ordenado de forma también cíclica con base en alguna variable o factor que influya en el criterio que queremos medir. Por ejemplo, supongamos que queremos medir la proporción de visitas a la que se le toma la tensión y que utilizamos como marco muestral las hojas de morbilidad o un registro similar de los pacientes atendidos cada día; supongamos además que algunos de los profesionales del centro tienen por costumbre tomar tensiones los miércoles. Si realizamos un muestreo sistemático sobre las hojas de consulta de –digamos– un mes, puestas en orden cronológico día tras día, corremos el riesgo de sobrestimar o subestimar la variable que queremos medir porque en el universo va a estar sistemáticamente agrupada en determinados días que pueden saltarse sistemáticamente o ser incluidos sistemáticamente en la muestra. En este caso el muestreo sistemático no sería aconsejable.

12. Muestreo estratificado. Metodología, ventajas e inconvenientes

El muestreo estratificado es aquel en el que la muestra (n) se subdivide en estratos o submuestras extraídas cada una de los diferentes estratos en que a su vez está dividido el universo o marco muestral. Los estratos del universo o marco muestral se definen en

Para el muestreo aleatorio simple, todos los elementos del universo han de estar numerados y ser localizables.

El muestreo sistemático es muy cómodo de realizar, sólo exige saber el total del universo y el número de casos que queremos. Con ello se calcula el intervalo de muestreo. Hay que estar alerta, en cambio, a los factores que intervienen en la forma en que está ordenado el universo para evitar sesgos.

En el muestreo estratificado la muestra se subdivide en submuestras que se extraen de los correspondientes estratos en que se ha dividido el universo. Los estratos son agrupaciones de los casos en función de la variable (por ejemplo, médicos) que queremos controlar con la estratificación.

función de las categorías o posibilidades de la variable o factor por el que nos interesa estratificar, el cual habitualmente influye en lo que queremos medir y, por lo tanto, nos interesa controlar su representatividad en la muestra.

En el ejemplo anterior, dado que sabemos que el día de la semana influye en la proporción de pacientes a los que se les toma la tensión, nos interesaría hacer un muestreo estratificado por días de la semana, de forma que dividiríamos la muestra en cinco submuestras (exceptuamos sábado y domingo) y cada una se extraería de entre los pacientes atendidos en cada uno de los cinco días de la semana en los que hay consulta normal.

Otra situación muy frecuente que aconseja la estratificación es la que ocurre cuando queremos que nuestro muestreo sea representativo de todo el centro o servicio, pero sabemos o sospechamos que puede haber grandes diferencias en el cumplimiento de los criterios entre los diferentes médicos o profesionales individuales implicados en el cumplimiento de los criterios. En este caso conviene también asegurarnos de la representatividad de la muestra, considerando a cada profesional individualizado como un estrato y realizando un muestreo estratificado.

La estratificación de la muestra puede realizarse de forma proporcional o no proporcional a la contribución de cada estrato al total del marco muestral. Para la estratificación *proporcional*, se ha de calcular primero en qué proporción contribuye cada estrato al total del universo o marco muestral. Por ejemplo, supongamos que queremos estratificar por médico, que en el centro trabajan cuatro médicos, y que el total de casos del universo es 500, pero que a estos 500 casos, el médico 1 contribuye con 100, el médico 2 con 50, el médico 3 con 200 y el médico 4 con 150. Supongamos además que queremos extraer una muestra de 60 casos y que la queremos estratificar de forma proporcional. Calcularemos entonces en qué proporción contribuye cada médico al universo:

$$\text{médico 1:} = \frac{100}{500} = 0.2 \rightarrow 20\%$$

$$\text{médico 2:} = \frac{50}{500} = 0.1 \rightarrow 10\%$$

$$\text{médico 3:} = \frac{200}{500} = 0.4 \rightarrow 40\%$$

$$\text{médico 4:} = \frac{150}{500} = 0.3 \rightarrow 30\%$$

La muestra se divide entonces en estratos proporcionales a estos mismos valores. Así:

de los pacientes del médico 1 se extrae 20% de 60 → 12 casos,
de los pacientes del médico 2 se extrae 10% de 60 → 6 casos,
de los pacientes del médico 3 se extrae 40% de 60 → 24 casos,
de los pacientes del médico 4 se extrae 30% de 60 → 18 casos.

En el muestreo estratificado no proporcional se extrae un número fijo o un porcentaje de la muestra fijo de cada estrato, con independencia del tamaño del estrato. El que se haga proporcional o no influye en los cálculos necesarios para valorar el resultado, como veremos en el capítulo 7.

A su vez, hay que extraer cada submuestra (la muestra de cada estrato) de forma aleatoria, sea simple o sistemática. El muestreo estratificado y sistemático es muy frecuente en los estudios del nivel de calidad.

Una cuestión común a todos los tipos de muestreo y que hay que tener prevista también en el diseño del estudio es el *mecanismo de sustitución de los casos* cuando éstos no se encuentran (por ejemplo, historias que no están) o no contienen datos válidos. El mecanismo de sustitución debe explicitarse de antemano y ser siempre el mismo en todo el muestreo. Una forma común es elegir el número o la historia que aparece en el listado inmediatamente después de la que ha fallado.

13. Muestreo de conveniencia. Metodología, ventajas e inconvenientes

El muestreo de conveniencia es de fácil obtención pero no garantiza la representatividad estadística. Se trata de una forma de muestreo en la que se eligen los casos que están más a la mano, para tener una idea de si hay o no pro-

El muestreo estratificado puede ser proporcional o no proporcional al tamaño de los estratos. En cada estrato el muestreo ha de ser aleatorio, simple o sistemático.

Los muestreos de conveniencia, aunque puedan aportar información de interés, generalmente no son estadísticamente representativos y hay que acompañarlos de una reflexión sobre el grado de generalización de lo que estamos midiendo.

blema; por ejemplo, si decidimos evaluar los pacientes que vengan hoy, o los que vinieron ayer, o la semana pasada, o las primeras veinte historias, etc. Normalmente se establece de antemano un número de casos también cómodo (muestreo tipo cuota) e incluso se aplican metodologías semejantes a la estratificación (muestréos *a propósito*) eligiendo cuotas de casos extraídos a conveniencia de grupos diferentes de pacientes (de diferentes médicos, por ejemplo) sobre los que se quiere saber si existen o no diferencias.

No es un muestreo de primera opción para los estudios de nivel de calidad, pero no es infrecuente que se realice este tipo de muestréos (por ejemplo, evaluar los casos atendidos en el último mes, o en los últimos seis meses). En estos casos hay que reflexionar hasta qué punto es o no una muestra representativa de lo que queremos medir; es decir, si hay suficientes elementos como para asumir que la forma de prestar el servicio que se reflejará en el muestreo de conveniencia va a representar la forma común o habitual de actuar. A veces, aunque parezca un muestreo de conveniencia sí es representativo, y al revés: muestréos que se suponen representativos son en realidad de conveniencia; esto depende de si el marco muestral escogido para el muestreo de conveniencia representa razonablemente la población o marco muestral más amplio al que queremos generalizar los resultados de la medición. Por ejemplo, si los pacientes de la última semana representan los que podían verse en cualquier semana del año, los resultados pueden ser generalizados a todo el año. La reflexión sobre la representatividad de la muestra (control del error de sesgo) debe de hacerse siempre en el momento del diseño del estudio.

14. Tipo de evaluación, según relación temporal con la acción evaluada

Si lo que se evalúa ha ocurrido ya, es una evaluación *retrospectiva*; si está ocurriendo a la vez que se evalúa, es *concurrente*. La mayoría de las evaluaciones del nivel de calidad son retrospectivas: se evalúa algo que ha ocurrido ya y, por lo tanto, si ha habido problema o defecto de calidad, es ya inevitable. La utilidad de

estas evaluaciones estriba en aprovechar esta información para diseñar intervenciones que eviten que el problema vuelva a ocurrir.

En las evaluaciones concurrentes, puesto que se realizan antes de que se dé por terminada la acción que se evalúa, es posible introducir las medidas correctoras antes de que se consume un acto asistencial de mala calidad. Por ejemplo, si evaluamos la calidad de la información que recibe el paciente sobre su tratamiento, con base en un diseño de encuesta realizada a los usuarios antes de que abandonaran el centro de salud o consultorio sometido a evaluación, y en la evaluación encontramos que el paciente no tiene claros algunos criterios empleados en la evaluación (dosis, pauta, contraindicaciones, tiempo de duración del tratamiento), entonces éstos se le aclaran y nos aseguramos de que los comprenda antes de que abandone la institución. En la asistencia hospitalaria hay más ejemplos de evaluación concurrente, algunos de ellos realizados de forma rutinaria aunque no siempre con criterios de evaluación explícitos. Citemos como ejemplos relativamente rutinarios la solicitud de segundas opiniones antes de proceder a determinada intervención o exploración, la práctica de aprendizaje supervisada por un docente que interviene en caso necesario modificando un curso de acción determinado que le parece erróneo, o la realización de biopsias intraoperatorias antes de proceder a la realización de determinadas intervenciones quirúrgicas. Las evaluaciones concurrentes, si se encuentra la ocasión y la oportunidad de ser realizadas, pueden convertirse con facilidad en parte integrante del proceso asistencial en sí.

El término de *evaluación prospectiva* (antes de que ocurra lo que se evalúa) es equívoco y preferimos no utilizarlo al hablar de estudio de nivel de calidad, puesto que más bien responde a actividades de diseño de la calidad, tal como vimos en el capítulo 1.

15. Tipo de evaluación, según quién toma la iniciativa

Existen dos tipos complementarios de organización de las actividades de GC: programas internos y programas externos. El esquema metodológico es en esencia el mismo, pero se

En el diseño de la evaluación hay que decir qué tipo de evaluación, retrospectiva o concurrente, vamos a realizar.

También hay que explicitar si la evaluación es interna, decidida por los profesionales evaluados, o externa.

diferencian en quienes toman la iniciativa de las evaluaciones y demás actividades del programa: en los programas internos, la responsabilidad de las actividades y la iniciativa de las mismas recae en los propios profesionales cuya asistencia se evalúa; en los programas externos, la responsabilidad de las actividades y la iniciativa de las mismas recae en organismos externos a los centros o profesionales evaluados. Estas instancias externas pueden ser niveles superiores de la organización administrativa del sistema de salud u organizaciones profesionales de carácter independiente.

Merece la pena aclarar que el hecho de que la evaluación sea interna no significa necesariamente que sean los propios profesionales del centro los que lleven a cabo la recogida física de datos: en determinadas situaciones estructurales favorables es perfectamente posible –y a veces deseable– que la recogida de datos la realice personal ajeno al centro, administrativo o contratado para éstas y otras tareas mayormente rutinarias. Una revisión interna se caracteriza por responder a un diseño y llevar a una explotación de los resultados para mejorar la asistencia, que responden a la iniciativa consciente y responsable de los profesionales de la institución.

16. Tipo de evaluación, según quién obtiene los datos

Ya hemos visto que la iniciativa de la evaluación puede ser interna o externa, y que este punto debe ser consignado en el diseño del estudio. Adicionalmente, hay que considerar y explicar quién se encarga de obtener los datos y de qué forma.

En las evaluaciones internas, es importante considerar si la evaluación será *cruzada* (un profesional evalúa la actuación de otro) o si será una *autoevaluación* (cada uno recoge los datos de su propia actuación). Esta segunda modalidad no es totalmente desdeñable en los estudios de calidad porque se supone que nadie está interesado en engañarse a sí mismo, sino en saber en qué debe mejorar; sin embargo, al menos en los primeros proyectos, quizá sea más aconsejable la evaluación cruzada o mixta. También puede ser realizada por personal ajeno al proceso evaluado.

Todo lo expuesto hasta aquí, toma cuerpo real y sentido al aplicarlo en la práctica. Como ayuda adicional, se incluye en este capítulo (anexo 5.3) un formato que resume lo que hay que describir en el proyecto de evaluación del nivel de calidad.

La evaluación puede realizarse de forma cruzada, por personal ajeno o evaluándose uno mismo los propios datos. El proceder concreto debe consignarse en el diseño del estudio.

BIBLIOGRAFÍA

- Argimon Pallás JM, Jiménez Villa J. Métodos de Investigación clínica y epidemiológica. 2a ed. Madrid: Harcourt-Brace, 2000.
- Briones G. Métodos y técnicas de investigación para las ciencias sociales. México: Trillas, 1990.
- Lemeshow S, Hosmer DW, Klar J, Lwanga SK. Adequacy of sample size in health studies. Sussex: WMO/John Wiley & Son, 1992.
- Palmer RH. Evaluación de la asistencia ambulatoria. Principios y práctica. Madrid: Ministerio de Sanidad y Consumo, 1990.
- Saturno PJ. Cómo diseñar una evaluación sobre la calidad asistencial en atención primaria. En: IV Congreso de la Sociedad Andaluza de Medicina Familiar y Comunitaria. Sevilla: Consejería de Salud, 1992.

Anexo 5.1

Número de casos necesarios para asegurarse un determinado intervalo (i) para una confianza de 95%, segúnsegún diversos valores esperados de la variable que se mide

VALOR ESPERADO (EN PROPORCIÓN) DE LA VARIABLE A MEDIR.

(i)	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
0,01	1825	3457	4898	6147	7203	8067	8740	9220	9508	9604	9508	9220	8740	8057	7203	6147	4898	3457	1825
0,02	456	864	1225	1537	1801	2017	2185	2305	2377	2401	2377	2305	2185	2017	1801	1537	1225	864	458
0,03	203	384	544	683	800	896	971	1024	1056	1067	1056	1024	971	896	800	683	544	384	203
0,04	114	216	306	384	450	504	546	576	594	600	594	576	546	504	450	384	306	216	114
0,05	73	138	198	246	288	323	350	369	380	384	380	369	350	323	288	246	198	138	73
0,06	51	96	136	171	200	224	243	256	264	267	264	256	243	224	200	171	136	96	51
0,07	37	71	100	125	147	165	178	188	194	196	194	188	178	165	147	125	100	71	37
0,08	29	54	77	96	113	126	137	144	149	150	149	144	137	126	113	96	77	54	29
0,09	23	43	60	76	89	100	108	114	117	119	117	114	108	100	89	76	60	43	23
0,10	18	35	49	61	72	81	87	92	95	96	95	92	87	81	72	61	49	35	18
0,11	15	29	40	51	60	67	72	76	78	79	78	76	72	67	60	51	40	29	15
0,12	13	24	34	43	50	56	61	64	66	67	66	64	61	56	50	43	34	24	13
0,13	11	20	29	36	43	48	52	55	56	57	56	55	52	48	43	36	29	20	11
0,14	9	18	25	31	37	41	45	47	49	49	49	47	45	41	37	31	25	18	9
0,15	8	15	22	27	32	36	39	41	42	43	42	41	39	36	32	27	22	15	8
0,20	5	9	12	15	18	20	22	23	24	24	24	23	22	20	18	15	12	9	5
0,25		6	8	10	12	13	14	15	15	15	15	15	14	13	12	10	8	6	

Adaptado de: Lemershow S, Hosmer OW, Kler J, Lwanga SK: Adequacy of Sample Size in Health Studies, HWO, 1990.

Anexo 5.2

Intervalo (i) para un grado de confianza del 95%, según estimación puntual y tamaño muestral

Estimación puntual (%)	Tamaño muestral (n)					Tamaño muestral (n)					
	30	50	60	100	200	30	50	60	100	200	
1	85,0	2,8	2,5	2,0	1,4	51	17,9	13,9	12,6	9,8	6,9
2	5,0	3,9	3,5	2,7	1,9	52	17,9	13,8	12,6	9,8	6,9
3	6,1	4,7	4,3	3,3	2,4	53	17,9	13,8	12,6	9,8	6,9
4	7,0	5,4	5,0	3,8	2,7	54	17,8	13,8	12,6	9,8	6,9
5	7,8	6,0	5,5	4,3	3,0	55	17,8	13,8	12,6	9,8	6,9
6	8,5	6,6	6,0	4,7	3,3	56	17,8	13,8	12,6	9,7	6,9
7	9,1	7,1	6,5	5,0	3,5	57	17,7	13,7	12,5	9,7	6,9
8	9,7	7,5	6,9	5,3	3,8	58	17,7	13,7	12,5	9,7	6,8
9	10,2	7,9	7,2	5,6	4,0	59	17,6	13,6	12,4	9,6	6,8
10	10,7	8,3	7,6	5,9	4,2	60	17,5	13,6	12,4	9,6	6,8
11	11,2	8,7	7,9	6,1	4,3	61	17,5	13,5	12,3	9,6	6,8
12	11,6	9,0	8,2	6,4	4,5	62	17,4	13,5	12,3	9,5	6,7
13	12,0	9,3	8,5	6,6	4,7	63	17,3	13,4	12,2	9,5	6,7
14	12,4	9,6	8,8	6,8	4,8	64	17,2	13,3	12,1	9,4	6,7
15	12,8	9,9	9,0	7,0	4,9	65	17,1	13,2	12,1	9,3	6,6
16	13,1	10,2	9,3	7,2	5,1	66	17,0	13,1	12,0	9,3	6,6
17	13,4	10,4	9,5	7,4	5,2	67	16,8	13,0	11,9	9,2	6,5
18	13,7	10,6	9,7	7,5	5,3	68	16,7	12,9	11,8	9,1	6,5
19	14,0	10,9	9,9	7,7	5,4	69	16,6	12,8	11,2	9,1	6,4
20	14,3	11,1	10,1	7,8	5,5	70	16,4	12,7	11,6	9,0	6,4
21	14,6	11,3	10,3	8,0	5,6	71	16,2	12,6	11,5	8,9	6,3
22	14,8	11,5	10,5	8,1	5,7	72	16,1	12,4	11,4	8,8	6,2
23	15,1	11,7	10,6	8,2	5,8	73	15,9	12,3	11,2	8,7	6,2
24	15,3	11,8	10,8	8,4	5,9	74	15,7	12,2	11,1	8,6	6,1
25	15,5	12,0	11,0	8,5	6,0	75	15,5	12,0	11,0	8,5	6,0
26	15,7	12,2	11,1	8,6	6,1	76	15,3	11,8	10,8	8,4	5,9
27	15,9	12,3	11,2	8,7	6,2	77	15,1	11,7	10,6	8,2	5,8
28	16,1	12,4	11,4	8,8	6,2	78	14,8	11,5	10,5	8,1	5,7
29	16,2	12,6	11,5	8,9	6,3	79	14,6	11,3	10,3	8,0	5,6
30	16,4	12,7	11,6	9,0	6,4	80	14,3	11,1	10,1	7,8	5,5
31	16,6	12,8	11,7	9,1	6,4	81	14,0	10,9	9,9	7,7	5,4
32	16,7	12,9	11,8	9,1	6,5	82	13,7	10,6	9,7	7,5	5,3
33	16,8	13,0	11,9	9,2	6,5	83	13,4	10,4	9,5	7,4	5,2
34	17,0	13,1	12,0	9,3	6,6	84	13,1	10,2	9,3	7,2	5,1
35	17,1	13,2	12,1	9,3	6,6	85	12,8	9,9	9,0	7,0	4,9
36	17,2	13,3	12,1	9,4	6,7	86	12,4	9,6	8,8	6,8	4,8
37	17,3	13,4	12,2	9,5	6,7	87	12,0	9,3	8,5	6,6	4,7
38	17,4	13,5	12,3	9,5	6,7	88	11,6	9,0	8,2	6,4	4,5
39	17,5	13,5	12,3	9,6	6,8	89	11,2	8,7	7,9	6,1	4,3
40	17,5	13,6	12,4	9,6	6,8	90	10,7	8,3	7,6	5,9	4,2
41	17,6	13,6	12,4	9,6	6,8	91	10,2	7,9	7,2	5,6	4,0
42	17,7	13,7	12,5	9,7	6,8	92	9,7	7,5	6,9	5,3	3,8
43	17,7	13,7	12,5	9,7	6,9	93	9,1	7,1	6,5	5,0	3,5
44	17,8	13,8	12,6	9,7	6,9	94	8,5	6,6	6,0	4,7	3,3
45	17,8	13,8	12,6	9,8	6,9	95	7,8	6,0	5,5	4,3	3,0
46	17,8	13,8	12,6	9,8	6,9	96	7,0	5,4	5,0	3,8	2,7
47	17,9	13,8	12,6	9,8	6,9	97	6,1	4,7	4,3	3,3	2,4
48	17,9	13,8	12,6	9,8	6,9	98	5,0	3,9	3,5	2,7	1,9
49	17,9	13,9	12,6	9,8	6,9	99	3,6	2,8	2,5	2,0	1,4
50	17,9	13,9	1,7	9,8	6,9	100	0,0	0,0	0,0	0,0	0,0

$$p \pm 1,96 \sqrt{\frac{p(1-p)}{n}}$$

p: estimación puntual
n: tamaño de la muestra

Si $\frac{n}{N} > 0,1$

$$i' = i \cdot \sqrt{\frac{N-n}{N-1}}$$

N: tamaño dl universo.

Anexo 5.3

Práctica recomendada. Formato para el diseño del estudio de nivel de calidad.

Se aconseja realizar el ejercicio de diseñar un estudio sobre nivel de calidad especificando cada uno de los componentes del esquema que se explica en este capítulo.

(Los espacios son exclusivamente orientativos)

TITULO DEL ESTUDIO:

PROBLEMA DETECTADO:

ANÁLISIS EFECTUADO: (Adjuntar diagrama de causa/efecto)

CRITERIOS PARA EVALUAR LA CALIDAD:

	<u>ENUNCIADO</u>	<u>EXCEPCIONES</u>	<u>ACLARACIONES</u>
CRITERIO 1:			
CRITERIO 2:			
CRITERIO 3:			
CRITERIO 4:			
CRITERIO 5:			
CRITERIO ...			

DIMENSIÓN ESTUDIADA:

TIPOS DE DATOS: (estructura, proceso y resultado)

Criterio 1:

Criterio 2:

Criterio 3:

Criterio 4:

Criterio 5:

Criterio ...

/... continuación

UNIDADES DE ESTUDIO (detallando si hay diferentes para cada criterio)

	<u>RECEPTORES DEL SERVICIO PACIENTES O UNIDADES DE ESTUDIO</u>	<u>PROVEEDORES</u>	<u>PROCESO EVALUADO</u>
CRITERIO 1:			
CRITERIO 2:			
CRITERIO 3:			
CRITERIO 4:			
CRITERIO 5:			
CRITERIO ...			

FUENTES DE DATOS

- Para la identificación de los casos o unidades de estudio:

- Para la obtención de los datos sobre cumplimiento de los criterios:

IDENTIFICACIÓN Y MUESTREO DE LOS CASOS

- Marco muestral:
- N° de casos a evaluar (tamaño de la muestra):
- Método de muestreo:
- Mecanismo de sustitución de casos:

TIPOS DE EVALUACIÓN

- En relación a la iniciativa para evaluar:
- En relación a la acción temporal con la acción evaluada:
- En relación a las personas responsables de extraer los datos:

Capítulo 6

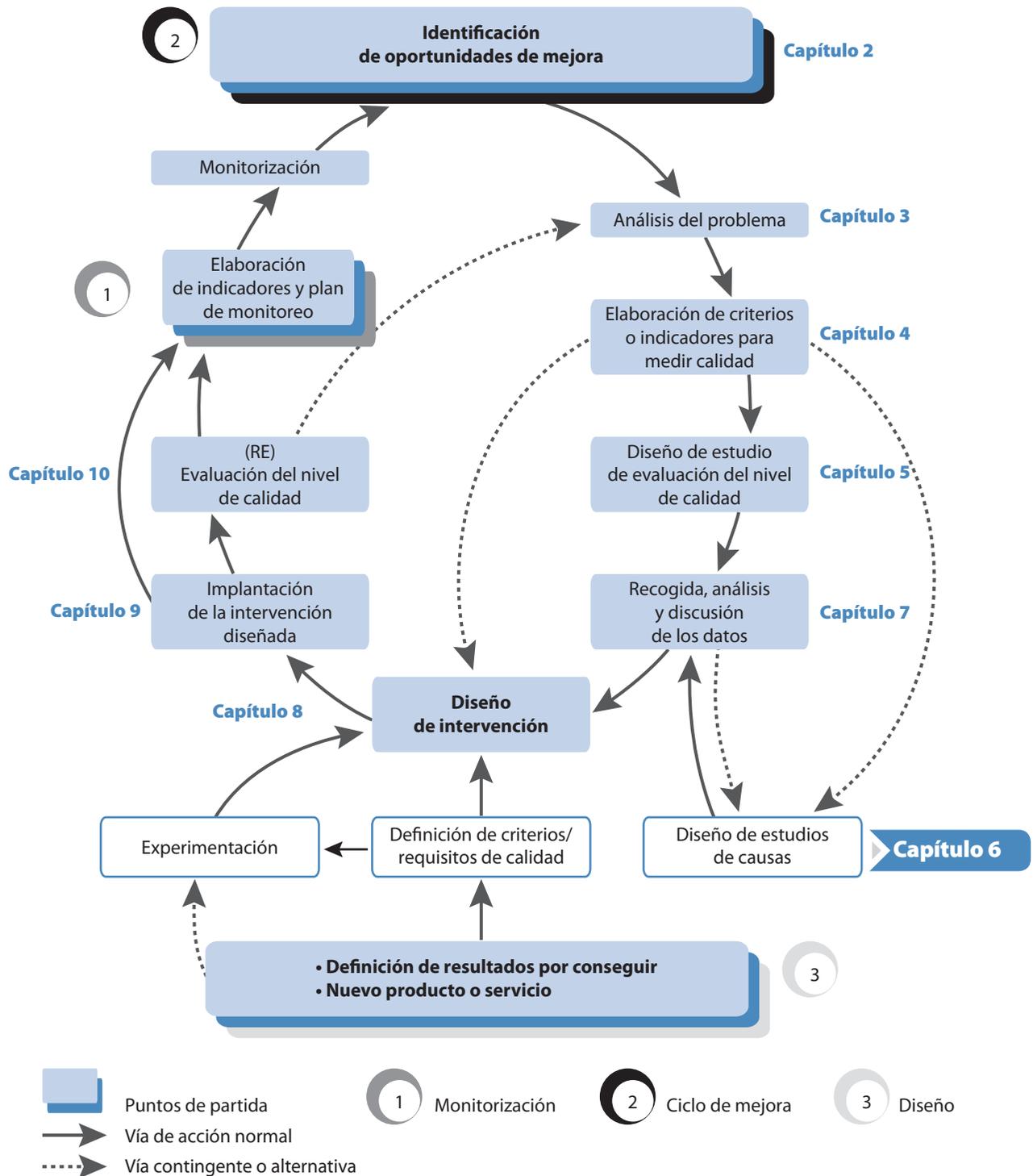
Diseño de estudios para la investigación de causas

Contenido general y objetivos específicos

1. Introducción
2. Métodos generales para los estudios de causas hipotéticas
3. Búsqueda de causas hipotéticas sin datos
4. Búsqueda de causas con datos: comprobación de hipótesis
5. Análisis estratificado
6. Correlación y diagrama de dispersión
7. Combinando métodos: las siete herramientas

Anexo 6.1. Indicaciones para la aplicación práctica de los contenidos de este capítulo

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Contenido general

Este capítulo complementa al anterior, detallando las características y metodología de los estudios de causas hipotéticas en los que no se miden una serie de requisitos o criterios de calidad, sino más bien uno solo, que representa el problema de calidad a estudiar y las causas que hipotéticamente pueden estar asociadas a su existencia. Se revisan métodos con y sin datos para identificar las causas sobre las que intervenir para mejorar. Para el estudio de causas con datos se indican los métodos, test estadísticos y herramientas gráficas que más conviene utilizar en función del tipo de variables (cuantitativa o cualitativa) con las que se miden las causas hipotéticas y los criterios o indicadores de calidad. El acercamiento metodológico de este capítulo es el adecuado para problemas organizacionales y de gestión, o de otro tipo, que precise del estudio de causas hipotéticas.

Objetivos específicos

En este capítulo se describen los métodos para:

1. Identificar el tipo de estudio que conviene para avanzar en el ciclo de mejora, según las causas a indagar.
2. Distinguir métodos de estudio de causas hipotéticas con y sin datos.
3. Diseñar y analizar una encuesta de causas.
4. Diseñar y analizar una votación múltiple sobre causas hipotéticas.
5. Identificar el tipo de variable que corresponde a los factores hipotéticos a indagar.
6. Seleccionar el test estadístico que corresponde a cada tipo de variables.
7. Describir el fundamento de los análisis de estratificación.
8. Describir el fundamento de los análisis de correlación y regresión.
9. Utilizar e interpretar adecuadamente un diagrama de dispersión.
10. Realizar la prueba de los signos en un diagrama de dispersión.

1. Introducción

Tras seleccionar la oportunidad de mejora o problema de calidad (capítulo 2) nos preguntamos de qué forma podemos abordarlo para conseguir su resolución. Para ello requerimos indagar sobre sus causas. En consecuencia, realizamos un análisis estructurado de la oportunidad de mejora (capítulo 3). Al realizar este análisis, lo más probable es que nos encontremos con diversos tipos de causas frente a las cuales la acción más oportuna es también diferente. Los dos cursos de acción a los que nos va a guiar el análisis del problema son la intervención directa y/o el diseño de estudios para un conocimiento cuantificado de las causas sobre las cuales basar la intervención.

Los estudios para un conocimiento cuantificado de las causas pueden definirse en dos grandes grupos de acuerdo a que las causas a indagar sean criterios o requisitos de calidad (estudios de nivel de calidad, capítulo 6) o bien factores hipotéticos que pueden o no relacionarse con el problema de calidad analizado (estudios de causas hipotéticas). Este último tipo de estudios es el contenido fundamental de esta unidad.

2. Métodos generales para los estudios de causas hipotéticas

Los estudios de causas hipotéticas son los que realizamos cuando el problema a analizar y mejorar no es de tipo exclusivamente clínico sino organizacional (tiempo de espera, descoordinaciones en el equipo, etc.), problema que podemos tener incluso ya medido, pero

que para su mejora tenemos que averiguar las causas que lo provocan; una situación parecida a la que tendríamos después de los estudios de nivel de calidad si no sabemos qué hacer para mejorar un determinado criterio o indicador clínico que no se cumple. En ambos casos, lo que tenemos es una serie de causas hipotéticas que investigar y es necesario medir concurrentemente con las posibles causas el criterio o los criterios que representan la calidad del aspecto en estudio. En ambos casos, también es necesario tener definido el criterio o criterios que nos miden el nivel de calidad del problema a analizar y mejorar.

En los estudios de causas hipotéticas, lo más conveniente sería diseñar estudios que las cuantifiquen. Sin embargo, en la práctica de la evaluación y mejora de la calidad en los servicios de salud se han aplicado también otros métodos que aprovechan de forma estructurada el conocimiento que pueden tener sobre el problema sus propios protagonistas. Son métodos como la encuesta de causas o la votación múltiple que parten o se complementan con los de análisis del problema y en los cuales se “cuantifican” las causas según la impresión subjetiva de los proveedores.

Cuando queremos que la cuantificación sea objetiva, podemos recurrir a un diseño de investigación que no se diferencia de los habituales estudios de evaluación de servicios de salud. Pero a veces procedimientos más simples, respetuosos pero con el rigor científico, pueden ser suficientes para el grado de conocimiento que necesitamos. De todos estos métodos de estudio (tanto con datos como sin datos), más propios del campo de la evaluación

En este capítulo nos vamos a ocupar de los estudios de causas hipotéticas de la mala calidad: aquellas que no son requisitos de calidad (criterios) pero que pensamos que pueden influir en que esos requisitos se den o no.

Las causas hipotéticas no son requisitos de calidad sino factores que pueden influir en ellos. Para estudiar si esta influencia es cierta hay que diseñar estudios que miden a la vez los requisitos de calidad (traducidos a criterios o indicadores) y las causas hipotéticas.

La valoración de la importancia de las causas hipotéticas puede hacerse de forma subjetiva (sin datos empíricos) o cuantificándolas objetivamente.

y mejora de la calidad, tratan los siguientes apartados de este capítulo.

3. Búsqueda de causas hipotéticas sin datos

Tras la identificación de posibles causas de un determinado problema de calidad por medio de métodos de consenso u otras como el diagrama de causa y efecto realizados sin datos, lo lógico sería diseñar un estudio que cuantificase su importancia relativa. Sin embargo, en algunos casos, por lo complicado del diseño, por la urgencia o la seguridad que puede tener el grupo en el conocimiento que tienen del problema, se puede proceder a señalar cuáles son las principales causas a remediar por medio de métodos que tratan de cuantificar ese conocimiento subjetivo. Para ello se realiza un listado de las causas hipotéticas según el resultado del análisis realizado y se someten a encuesta o a votación múltiple.

3.2. Encuesta sobre causas

En este enfoque, el listado de causas hipotéticas se convierte en un listado de preguntas, semejante a lo que se realiza en las encuestas de opinión, pidiendo al personal implicado que valore la importancia de cada una de ellas, con escalas de varias opciones (“no importante”, “poco importante”, “importante” o con más opciones).

Tanto el formato de las preguntas como la escala de valoración pueden ser de varios tipos, pero lo que se persigue es tener una cuantificación/priorización de las diversas causas en relación a su importancia. El resultado de la encuesta puede presentarse y analizarse con un diagrama de Pareto (explicado en detalle en el capítulo 7), en donde se subrayen y prioricen las causas sobre las que diseñar la intervención en función de la valoración recibida.

Este método ha sido empleado en algunos de los proyectos pioneros de aplicación a sistemas de salud de los métodos de evaluación y mejora de la calidad propios de la industria.

La figura 6.1 contiene un ejemplo real de resultado de encuesta de causas del retraso intrahospitalario en la aplicación de trombolíticos en el infarto de miocardio. En este caso, la

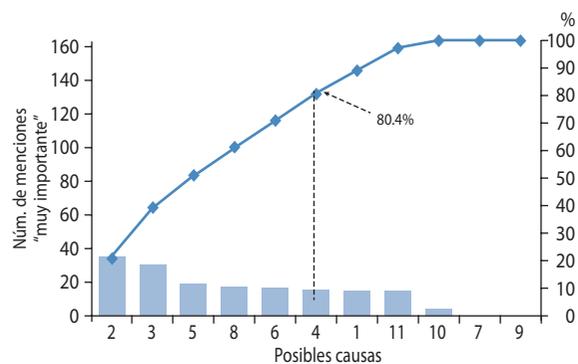
La encuesta de causas y la votación múltiple son los dos métodos más comunes para cuantificar la valoración subjetiva de las causas hipotéticas.

La encuesta sobre causas pide al personal implicado en el problema a evaluar que señale la importancia relativa de cada una de las posibles causas, sobre la base de su conocimiento directo del problema.

La encuesta se pasó a personal de Urgencias (n=65), pidiendo su valoración en una escala de tres opciones*

Listado de posibles causas de demora (en el orden en que figuran en la encuesta)

1. Pocos recursos humanos para el traslado del paciente hasta la UCI
2. Tiempo de preparación del paciente para el traslado (bala de oxígeno, desfibrilador, etc.)
3. Inadecuada estructura física de comunicación Urgencias-UCI
4. Interpretación del ECG por residentes
5. Necesidad de interconsulta residentes/staff
6. Demora por evaluación de pruebas complementarias (CPK, etc.)
7. Inadecuado sistema de telefonía interna para comunicación Urgencias-UCI
8. Demora por realización de diversas técnicas o exploraciones (Rx, vía venosa, etc.)
9. Demora por no priorización
10. Demora en el ingreso en UCI por ocupación
11. Otras causas



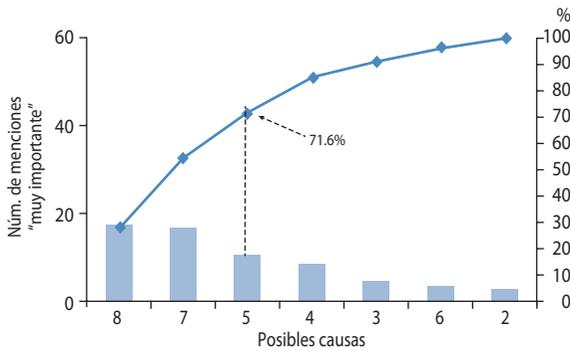
Posibles causas de demora	Núm. de menciones como "muy importante"	%	% acumulado
2	35	21.3	21.3
3	30	18.3	39.6
5	19	11.6	51.2
8	17	10.4	61.5
6	16	9.8	71.3
4	15	9.1	80.4
1	14	8.5	89.0
11	14	8.5	97.5
10	4	2.4	100.0
7	0	0.0	100.0
9	0	0.0	100.0
Total	164	100.0	

Figura 6.1. Encuesta de causas de retraso en la aplicación de trombolíticos

cuantificación, aunque subjetiva, fue acertada y la intervención diseñada sobre esta base fue efectiva. La figura 6.2 contiene un ejemplo de

Listado de posibles causas de no disponer de la historia clínica en la consulta

1. No tiene historia clínica
2. No se identifica bien de qué profesional son las historias
3. Retraso en la comunicación del cambio de médico al personal administrativo
4. La historia no está archivada
5. La historia está siendo utilizada por otro profesional en otra consulta
6. La historia está siendo utilizada para trabajos de investigación
7. La historia no se encuentra por estar incorrectamente archivada
8. El nombre de la cita y el de la historia no coinciden



Posibles causas de demora	Núm. de menciones como "muy importante"	%	% acumulado
8	17	28.3	28.3
7	16	26.7	55.0
5	10	16.7	71.6
4	8	13.3	85.0
3	4	6.7	91.6
6	3	5.0	96.6
2	2	3.3	100.0
Total	60	100.0	

Figura 6.2. Encuesta de causas de no tener la historia clínica en la consulta

cuantificación mediante encuesta, también real, de las causas de no tener la historia clínica en la consulta que no resultó acertada, a tenor del escaso efecto de la intervención diseñada con estos datos.

3.2. Votación múltiple

La votación múltiple es también un método para seleccionar las causas más importantes de

la lista que se haya elaborado. Consiste en una serie de votaciones entre los implicados en el proceso a mejorar, cada una de las cuales reduce la lista, hasta que quedan seleccionadas las causas sobre las que intervenir.

Los pasos a seguir son los siguientes:

1. Preparar el listado de causas.
2. Tras examinarlo, cada persona implicada entrega una lista en la que ha seleccionado un tercio de las causas de la lista (por ejemplo, si el listado tiene 30 causas, selecciona 10; si tiene 15, selecciona 5, etc.).
3. Se cuentan los "votos", viendo el número de menciones para cada causa de la lista, y se eliminan aquellas con menos menciones, reduciendo la lista. El número mínimo de votos para mantener la causa en la lista depende del tamaño del grupo que vota; la regla general es eliminar aquellas que tienen un número de votos inferior a 1/3 del tamaño del grupo.
4. Se repiten los pasos 1 a 3 hasta que el listado sea lo suficientemente reducido e importante como para ser objeto de diseño de intervención para contrarrestar las causas que contiene.

La votación múltiple busca el consenso de los implicados en el problema evaluado sobre cuáles son las causas principales, a través de un proceso estructurado de votación y reducción consecuente del listado de causas inicial.

La votación múltiple no deja de ser una técnica de consenso que tiene el riesgo, como todas las técnicas que no manejan datos, de equivocarse si las percepciones del grupo no son las correctas. La reevaluación de la calidad con base en el criterio/indicador que representa el problema, y su comparación con la situación (medición) de partida va a ser la única forma de saber si el grupo estuvo acertado o no.

Todos estos métodos, al igual que otras técnicas de consenso como el grupo nominal (capítulo 2) o el diagrama de afinidades (capítulo 8), pueden utilizarse tanto para priorizar causas como para diseñar la estrategia o intervención para mejorar; corriendo eso sí, el mismo riesgo de equivocarse.

4. Búsqueda de causas con datos: comprobación de hipótesis

Los componentes de los estudios para cuantificar las posibles causas de un problema de calidad y averiguar su asociación con el criterio o indicador con el que lo medimos no son

El diseño de los estudios para cuantificar las causas hipotéticas es idéntico al de los estudios de nivel de calidad; si bien en cada caso hay que medir tanto el criterio o indicador de calidad como las causas en consideración. La mayor diferencia entre los dos tipos de estudios radica en el análisis estadístico de los datos.

diferenciables en cuanto a su diseño de los que pueden ponerse en marcha para evaluar el nivel de calidad (capítulo 5). La diferencia es que en la muestra de casos en las que evaluamos la calidad (normalmente mediante un solo criterio o indicador que define el problema de calidad, por ejemplo retraso en la fibrinólisis o % de consultas sin historia, en los ejemplos de las figuras 6.1 y 6.2) se miden también las causas, características o factores que sospechamos son de interés. Con estos datos lo normal sería proceder a valorar la asociación entre el indicador de calidad y sus posibles causas; para ello, según el esquema de la figura 6.3, disponemos de una serie de test estadísticos (correlación de Pearson, correlación de Spearman, ANOVA, χ^2) a elegir y utilizar según la naturaleza esta-

dística de las dos variables (indicador de calidad y causa) cuya asociación queremos ver. Si quisiéramos considerar simultáneamente más de una posible causa, utilizaríamos entonces técnicas de análisis multivariable, sin embargo, en la práctica normal de los ciclos de mejora es frecuente no llegar a utilizar estos análisis estadísticos un tanto más complejos.

En la práctica, los estudios de causas hipotéticas pueden resumirse en tres tipos de situaciones y los análisis en dos enfoques principales: estratificación y correlación, cada uno de ellos adecuado según el tipo de variables de las que se trate. Tanto la estratificación como la correlación pueden acompañarse o no de análisis de significación estadística, utilizando entonces el test que resulte más apropiado según

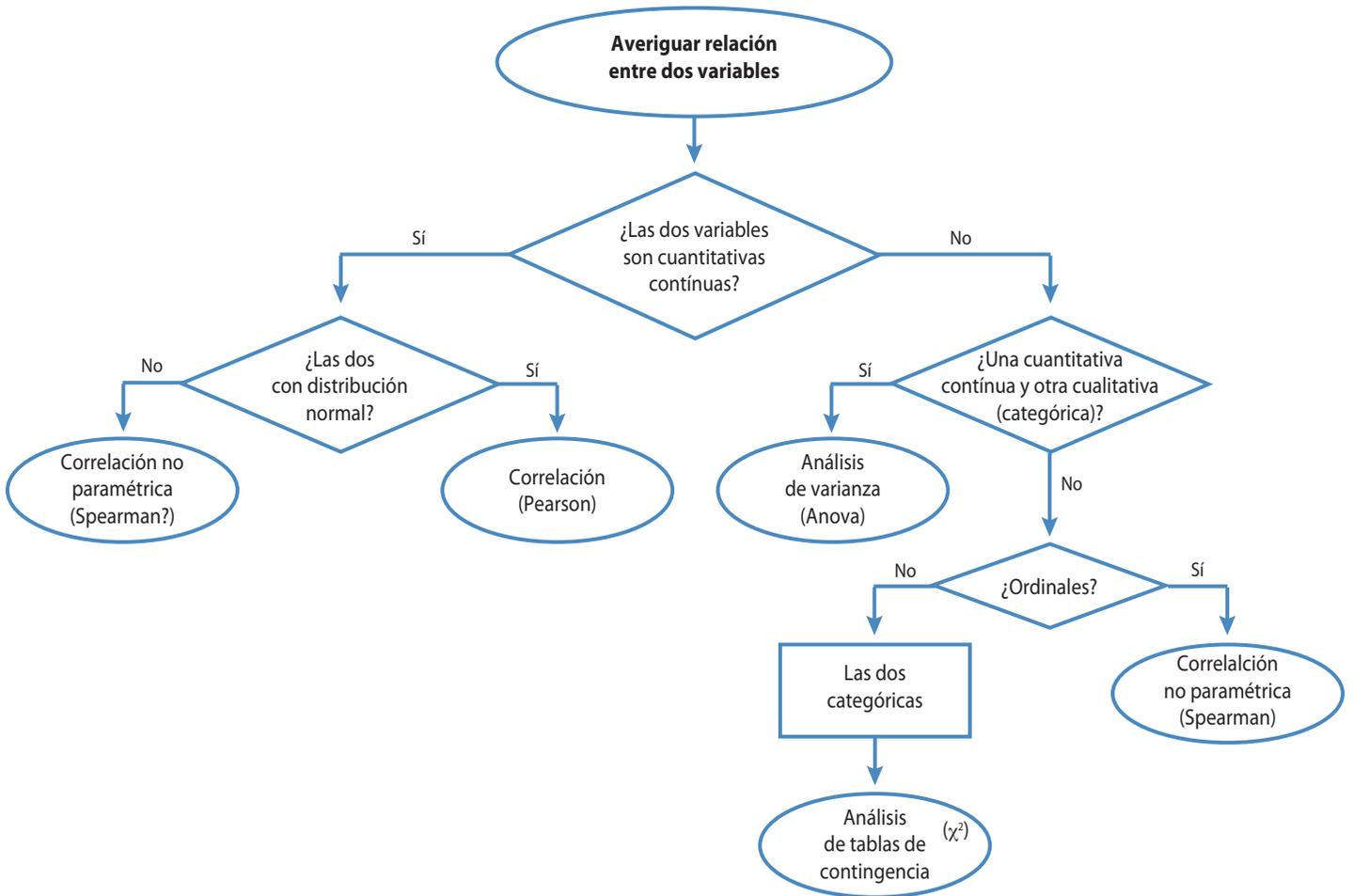


Figura 6.3. Test estadísticos para asociación entre dos variables (Indicador de calidad y una probable causa)

el algoritmo de la figura 6.3. Las tres posibles situaciones que podemos tener se definen en función de la caracterización de las causas a investigar y son las siguientes:

(I) Las causas hipotéticas son una serie o *conjunto de características*, circunstancias o factores de carácter cualitativo, dicotómico, de forma que mediremos su presencia o ausencia de manera similar a como lo haríamos si fuese un listado de criterios de calidad. Por ejemplo, supongamos que por medio de un flujograma como el de la figura 6.4 hemos identificado una serie de puntos problemáticos en el proceso de petición, realización y remisión de analítica que pueden provocar vueltas atrás al flujo normal de actividades y, por tanto, causar repeticiones y retrasos; en nuestro estudio de causas mediremos la frecuencia de cada una de estas posibles vueltas atrás (presencia/ausencia) cuando se ha producido un retraso problemático (previamente definido como problema de calidad). Para estar más seguros podemos medir si estas vueltas a atrás se producen también en casos que no se atrasan y comparar las frecuencias en ambos grupos. Adicionalmente, si el flujograma es correcto y todos los puntos problemáticos son indeseables, su presencia puede ser considerada como defectos de calidad y analizados con diagrama de Pareto (descrito en detalle en el capítulo 7) para priorizar las intervenciones para mejorar. Un caso semejante sería el análisis a efectuar sobre las causas de insatisfacción a través de encuesta; cada posible causa se traduce en una pregunta que es un factor o variable que puede considerarse como satisfactoria o insatisfactoria y realizar un análisis con diagrama de Pareto para evidenciar cuáles son los factores/causa más frecuentemente valorados como insatisfactorios.

(ii) La causa o factor a indagar es una *variable cualitativa*, categórica, con dos o más valores. Los casos más típicos son tipos de centros médicos diferentes dentro de un equipo, turnos, tipo de consulta, tipo de especialista, días de la semana, etc. El análisis oportuno en estos casos es la *estratificación* o análisis estratificado del indicador de calidad según los diversos valores del factor o causa a indagar. Este método se explica más adelante en este capítulo.

(iii) La causa o factor a indagar es una variable cuantitativa porque se mide así en cada individuo de la muestra (por ejemplo el tiempo de espera, n° de visitas, etc.) o porque las unidades de estudio son agrupaciones de individuos y lo que se mide es la proporción de una determinada cualidad (por ejemplo la proporción de consultas a demanda en un día, siendo el día la unidad de estudio tanto para el indicador de calidad como para el posible factor o causa asociada). En estos casos se analizará la correlación con el indicador de calidad si este es de tipo cuantitativo, o se realizará análisis estratificado de la probable causa según las dos posibilidades (cumple/no cumple) del indicador de calidad, si este es de tipo cualitativo. La correlación puede analizarse gráficamente por medio de un *diagrama de dispersión* (*Scatter Diagram*).

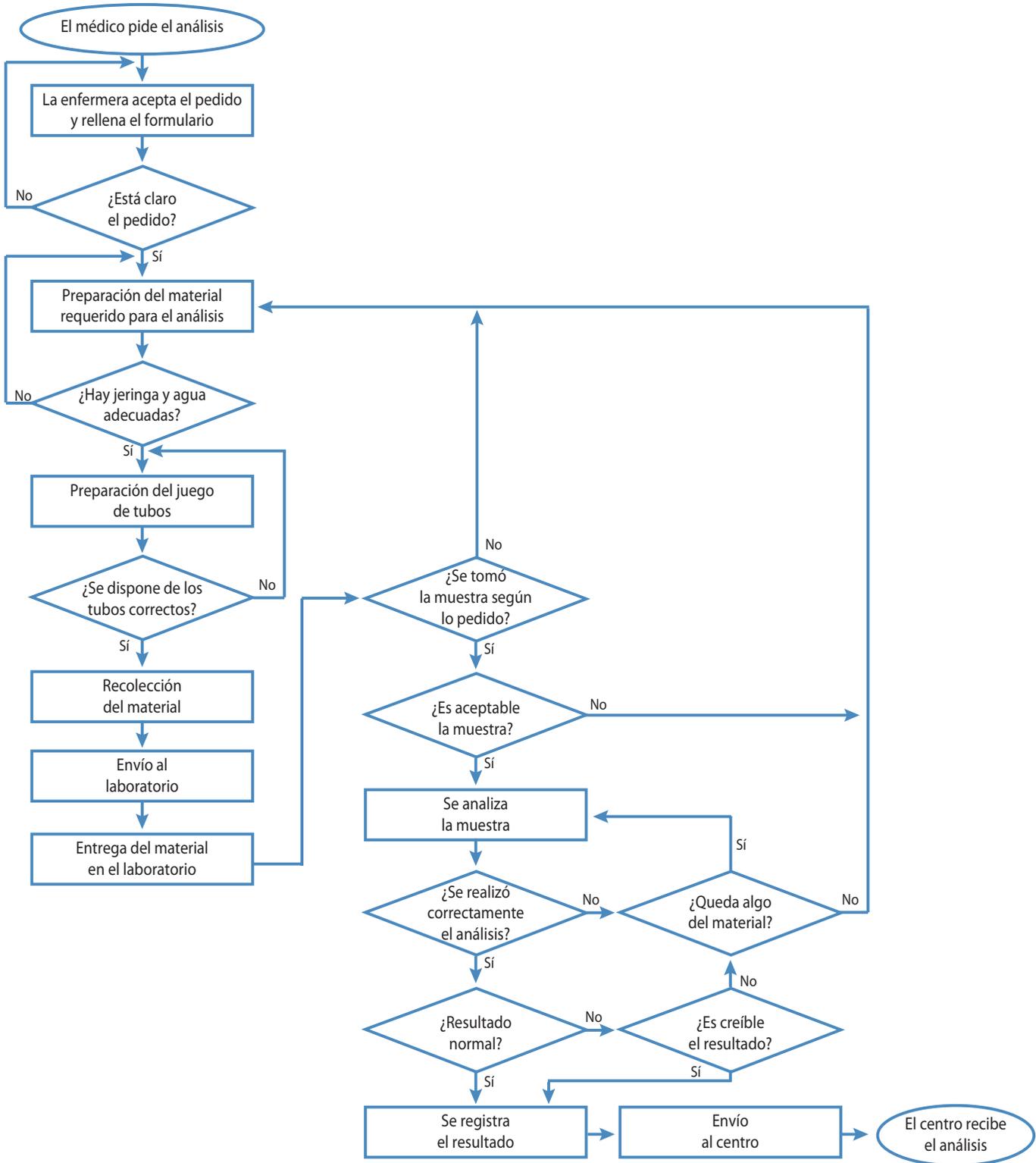
Es conveniente señalar también que todos estos tipos de análisis pueden combinarse; de hecho, en ocasiones combinarlos es lo más adecuado a medida que queremos profundizar en el análisis de las posibles causas. Por ejemplo un análisis de correlación puede ir seguido o acompañado de análisis de correlación estratificado, si la reflexión sobre las causas así lo aconseja porque la considerada en primer lugar resume una probable variabilidad interna que encubre la verdadera causa (por ejemplo: una media diaria puede encubrir diferencias importantes por horas, turnos, médicos, etc.).

Vamos a ver con más detalle en qué consiste la estratificación (útil para la situación (ii) descrita más arriba) y el análisis con diagrama de dispersión (útil para la situación (iii)).

5. Análisis estratificado

La estratificación puede ser definida como un método para identificar la fuente o causa de variación en los criterios o indicadores de calidad *clasificando las mediciones según las categorías o posibilidades (estratos) del factor a indagar*; un factor que pensamos puede causar diferencias en el indicador de calidad. De esta forma se describe de manera sistemática el cumplimiento o los valores del criterio o indicador en cada uno de los estratos, con lo que obtendremos una

Quando el criterio de calidad que representa el problema es dicotómico (cumple/no cumple) y las causas hipotéticas son asimismo un listado de características dicotómicas (presencia/ausencia) su análisis puede realizarse con un diagrama de Pareto de las veces que cada causa ha estado presente en casos con defecto de calidad.



Conclusión: Identificadas las posibles causas del retraso (bucles hacia atrás) diseñamos un estudio para medir cómo es de frecuente cada una y si se asocian o no al problema analizado (no tener a tiempo el análisis).

Figura 6.4. Flujograma del circuito de petición y realización de análisis para identificar causas de retraso

comprensión más detallada de la estructura y funcionamiento del proceso analizado. Un *estrato* puede ser definido como un segmento o categoría que se diferenciará, probablemente, del resto de categorías con respecto a la característica de calidad que se está analizando.

Supongamos que nos interesa mantener un balance entre las consultas a demanda y las programadas, y elegimos como indicador para medir este balance el porcentaje de consultas a demanda sobre el total de consultas o visitas del día. Tenemos la sensación de que son pocas las consultas programadas y que actuamos sólo a demanda. Para valorar el nivel de este indicador hemos recogido los datos diarios de las últimas cinco semanas y nos ha resultado una media de 80.5% y los valores diarios que figuran en el cuadro 6.1.

Una primera forma de analizar los datos sería ver si la media es estable o existe una gran dispersión, lo que indicaría; o bien un patrón rígido, insensible a las diferencias o altibajos que puede tener la demanda día a día, o bien la inexistencia de un patrón de conducta establecido. Este análisis puede hacerse con un gráfico de control (una metodología útil tanto para analizar problemas como para monitorizarlos) que en nuestro ejemplo ha resultado ser como el que se muestra en la figura 6.5. En él puede verse que la proporción de consulta a demanda es muy variable; con grandes altibajos y diferencias con respecto del valor medio de la serie. Para intentar explicar esta variabilidad podemos pensar en varios *factores o causas hipotéticas* como: el día de la semana, porque sabemos que la demanda varía por días, el

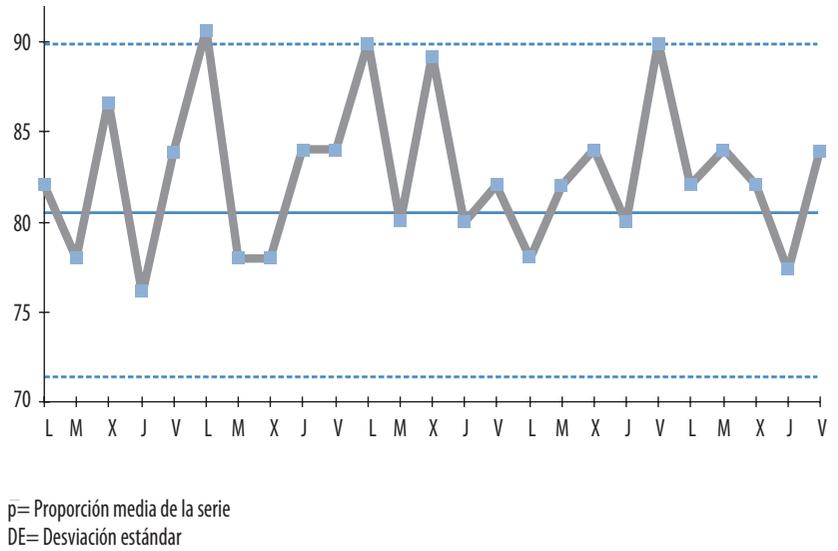


Figura 6.5. Porcentaje de consultas a demanda sobre total consultas. Evolución de los resultados diarios en las últimas cinco semanas. Gráfico de control

propio número diario de consultas a demanda (la hipótesis sería que mantenemos un número adecuado diario de consultas programadas, pero las diferencias se dan por las fluctuaciones de la demanda, que no conocemos bien y no tenemos en cuenta), etc.

Si nos fijamos primero en el primer factor mencionado, influencia del día de la semana, haríamos una estratificación de los datos según el día de la semana (con el resultado que puede verse en el cuadro 6.2, apuntando a una mayor demanda en lunes y viernes), y procederemos a comprobar si hay o no diferencia significativa entre ellos (por ejemplo, con una

Cuadro 6.1
Porcentaje de consultas a demanda sobre total consultas. Resultados diarios en las últimas cinco semanas

%		%		%		%		%
82.0	Lunes	90.7	Lunes	90.0	Lunes	78.0	Lunes	82.0
78.0	Martes	78.0	Martes	80.0	Martes	82.0	Martes	84.0
86.7	Miércoles	78.0	Miércoles	89.3	Miércoles	84.0	Miércoles	82.0
76.0	Jueves	84.0	Jueves	80.0	Jueves	80.0	Jueves	77.3
84.0	Viernes	84.0	Viernes	82.0	Viernes	90.0	Viernes	84.0

\bar{p} =80.5

La estratificación del nivel de calidad según los diversos factores o causas hipotéticas nos puede indicar cuál de ellas se asocia a diferencias significativas en el nivel de calidad. Sin embargo la utilidad del análisis va a depender en primer lugar de una correcta selección o identificación de las causas a analizar.

Cuadro 6.2	
Porcentaje de consultas a demanda sobre total consultas. Análisis estratificado según días de la semana	
	Porcentaje de consultas a demanda
Lunes	85.1
Martes	80.2
Miércoles	83.0
Jueves	79.1
Viernes	84,5

$$\bar{p}=80.5$$

χ^2). También podríamos hacer un análisis gráfico, aprovechando que ya tenemos nuestro gráfico de control, si bien su comprensión precisa del conocimiento y manejo de este tipo de gráficos (explicados en detalle como método de monitorización).

Podemos seguir un procedimiento similar con cualquier otro factor (médicos, enfermeras, fisioterapeutas, tipo de centro, etc.) que nos parezca de interés, siempre que éste sea una variable categórica (de varias opciones). Si el factor o causa de interés es cuantitativa, como sería en nuestro ejemplo en vez del día de la semana, la influencia de la fluctuación en el número de consultas a demanda, el análisis lo haríamos por medio de correlación, y, a nivel gráfico, con un diagrama de dispersión, tal como veremos más adelante.

Cuál de los factores por los que queremos estratificar va a mostrarse relevante es algo que no se sabe hasta que se realiza el análisis. En los manuales sobre control de calidad se encuentran ejemplos a veces tan llamativos como el análisis provocado por un incremento inusual de los accidentes de tráfico en Japón en 1970; como en los impresos sobre los accidentes había un apartado llamado “causa” se realizó un primer análisis viendo cuál era la más frecuente; ésta resultó ser “conducción imprudente”. Sin embargo, estratificando cada “causa” por las características de la carretera y la localización específica de los accidentes, se evidenciaron frecuencias significativamente más elevadas según determinadas localizaciones y condiciones de las carreteras: las verda-

deras causas sobre las que el Departamento de Carreteras decidió actuar, mostrándose después una drástica reducción en el número de accidentes de tráfico.

6. Correlación y diagrama de dispersión

La correlación es un análisis estadístico para comprobar la asociación entre dos variables de tipo cuantitativo. El resultado se expresa por medio de un coeficiente llamado *coeficiente de correlación* (r), que puede ir en valores absolutos de 0 a 1; expresando 1 el grado máximo de correlación, que si es positiva (cuando una variable aumenta, la otra también) se traduce en valores también positivos para el coeficiente de correlación, y si es negativa (cuando una variable aumenta, la otra disminuye) dará valores negativos para el coeficiente de correlación. La *regresión* es un análisis semejante, pero en el que estamos interesados no sólo en la existencia o no de asociación sino también en cuantificar en qué medida la diversificación de una variable (que llamamos independiente) provoca variación en la otra (que se llama dependiente), para lo cual se construye un modelo o ecuación que represente esa relación de dependencia y que puede llevarse a un gráfico por medio de la línea que resulte al aplicar a la ecuación las parejas de valores que hemos obtenido para las dos variables en estudio. Este acercamiento tiene una intención predictiva, (encontrar la fórmula con la cual pueden predecirse los valores de la variable dependiente en función de los que tenga la variable independiente), mientras que en la correlación solamente estamos interesados por la existencia de asociación; como si ambas variables fuesen variables “dependientes” entre sí.

En los análisis que se realizan en los ciclos de mejora es más frecuente el enfoque de correlación que el de regresión, pero en ambos casos, el análisis puede plasmarse e incluso realizarse de forma gráfica. De hecho, es un consejo muy extendido y aceptado que antes de aplicar ningún cálculo estadístico se debe observar la distribución gráfica de las mediciones.

Al plasmar nuestras mediciones en un gráfico (llamado gráfico de correlación o gráfico de dispersión), utilizamos un eje de coordenadas, siendo lo habitual el que en el eje de orde-

La correlación o asociación entre un criterio de calidad cuantitativo (por ejemplo tiempo de espera) y una causa hipotética también cuantitativa (por ejemplo número de pacientes de consulta a demanda) se puede valorar mediante un análisis estadístico (coeficiente de correlación) y también mediante un diagrama de dispersión, que puede adicionalmente analizarse calculando el coeficiente de correlación y su significación estadística, o mediante la prueba de los signos.

nadas (y) se represente el criterio o indicador de calidad que estamos analizando como problema, y en el de abscisas (x) la causa o factor influyente potencial; cada pareja de datos (valor del indicador y valor de la causa en estu-

dio) se representa como un punto en el gráfico. Como resultado podemos encontrarnos, en general, con los patrones que se reproducen en la figura 6.6. Si queremos ser más detallados en el análisis, puede calcularse el coeficiente de

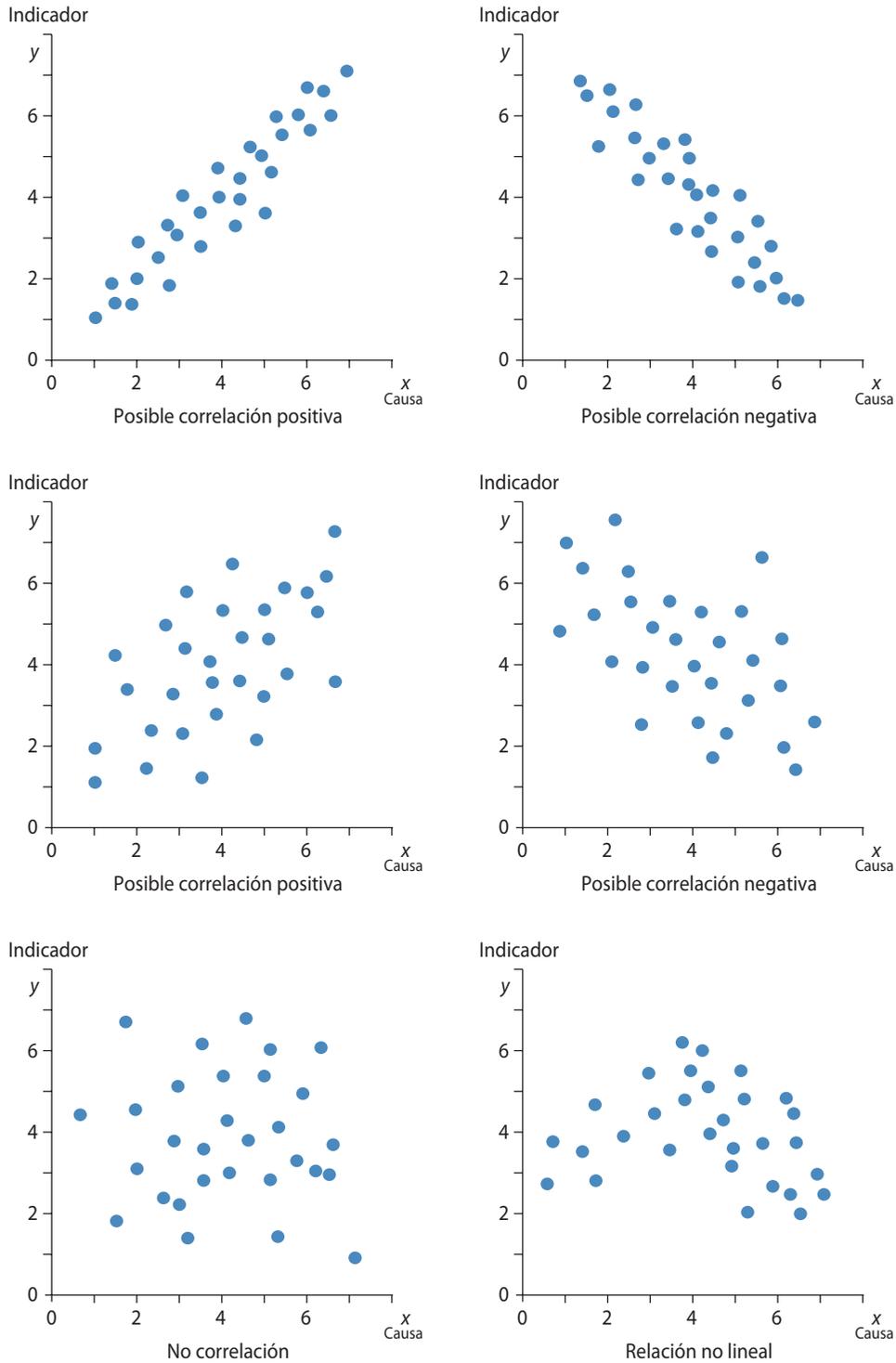


Figura 6.6. Posibles patrones de los gráficos de dispersión

correlación y su correspondiente significación estadística, o bien realizar una prueba no paramétrica sencilla basada en el simple conteo de los puntos en referencia a su posición en el gráfico (prueba de los signos).

6.1. Análisis de la correlación: prueba de los signos

Supongamos que nuestro indicador de calidad es la proporción de pacientes que esperan más de 45 minutos para ser atendidos y que pensamos que a pesar de nuestros esfuerzos por mantenerlo bajo, la propia variabilidad de la demanda influye en que nuestro indicador varíe, de forma que si hay más demanda, la proporción de pacientes que tienen que esperar más se eleva. Si esto es así, debemos intentar controlar mejor la demanda o establecer mecanismos organizativos que tengan en cuenta sus fluctuaciones para que no influyan en el tiempo de espera de los pacientes. Para estudiar esta relación entre tiempo de espera y porcentaje de consultas a demanda, recogemos los datos de ambas variables cada día durante un cierto tiempo (queremos tener al menos 30 mediciones) con el resultado que figura en el cuadro 6.3.

El correspondiente diagrama de dispersión es la figura 6.7, en el cual parece que, efectivamente, hay una relación positiva. Pero, ¿cómo podemos saber si es o no significativa? Por una parte, podemos calcular el coeficiente de correlación y ver su significación estadística, lo cual nos lo hará el ordenador con cualquier paquete estadístico u hoja de cálculo de que dispongamos, pero también podemos realizar este análisis visualmente si no disponemos de esos recursos o si, como ocurre con frecuencia, tenemos un número pequeño de mediciones. Para ello seguimos los siguientes pasos:

1. Trazamos una línea vertical a nivel de la mediana de los valores de la causa potencial (representada en el eje de abscisas, x), de forma que deja a cada lado la mitad de los valores para esta variable en nuestra serie. Para nuestro ejemplo, la mediana de esta causa potencial (el porcentaje de consultas a demanda) es 89.1.
2. Trazamos una línea horizontal a nivel de la mediana de valores del indicador anali-

Cuadro 6.3			
Porcentaje de pacientes que esperan más de 45 min (y), y porcentaje de consultas a demanda (x)			
	Fecha	y	x
Mayo	1	8.6	88.9
	2	8.9	88.4
	3	8.8	87.4
	4	8.8	89.1
	5	8.4	87.4
	8	8.7	88.6
	9	9.2	91.1
	10	8.6	91.2
	11	9.2	89.5
	12	8.7	89.6
	15	8.4	89.4
	16	8.2	86.4
	17	9.2	92.2
	18	8.7	90.9
	19	9.4	90.5
	22	8.7	89.2
	23	8.5	87.7
	24	9.2	88.5
	25	8.5	86.6
	26	8.3	89.6
	29	8.7	89.6
30	9.3	92.8	
31	8.9	88.6	
Junio	1	8.9	90.8
	2	8.3	88.1
	5	8.7	88.2
	6	8.9	90.4
	7	8.7	91.2
	8	9.1	92.5
	9	8.7	87.2

zado. En nuestro ejemplo, el indicador de calidad es la proporción de pacientes que esperan más de 45 minutos, la mediana es 8.7, valor que deja la mitad de valores por encima y la mitad por debajo.

3. Al trazar estas líneas tendremos el gráfico dividido en cuatro cuadrantes (n_1 , n_2 , n_3 , n_4). Procedemos entonces a contar el número de puntos en cada cuadrante, y su-

mamos los que hay en los cuadrantes contrapuestos (n_1+n_3 y n_2+n_4 , ver figura 6.7). En estos cálculos no se debe incluir ningún punto que esté justo en cualquiera de las dos medianas.

- La menor de estas dos sumas se compara con los valores del cuadro que se reproduce como cuadro 6.4, y si este valor es *igual o menor* que el que aparece en el cuadro para el correspondiente número total de puntos del gráfico, concluiremos que la correlación es significativa. Adicionalmente, si n_1+n_3 es mayor que n_2+n_4 la correlación es positiva, mientras que si n_1+n_3 es menor que n_2+n_4 la correlación es negativa, aunque esto es visible a simple vista en el gráfico.

En nuestro ejemplo, sin contar los puntos que coinciden con cualquiera de las dos medianas, tenemos:

$$n_1=6 \quad n_2=4 \quad n_3=8 \quad n_4=3$$

$$n_1+n_3=14$$

$$n_2+n_4=7$$

La suma más pequeña es 7. Buscamos en el cuadro 6.4 el valor que corresponde para $n=30$, y como 7 es igual o menor (en este caso menor) que 9, que es el valor correspondiente al nivel de significación ≤ 0.05 (e incluso igual al correspondiente a significación $\leq 0,01$) concluiremos que ambas variables están significativamente correlacionadas.

Es preciso recordar, sin embargo, que encontrar correlación entre dos variables no quiere decir que la relación sea causal; puede que la relación se establezca a través de terceros factores y, en cualquier caso, la relación de causalidad va a precisar de otro tipo de estudios; entre otras cosas, el resultado que obtengamos en la reevaluación una vez hayamos actuado sobre la causa potencial, nos confirmará o no si la relación era causal.

7. Combinando métodos: las siete herramientas

En la búsqueda de causas de los problemas de calidad, el "viaje diagnóstico" según algunos autores o "recorrido diagnóstico" según otros, no hay recetas mágicas. Es frecuente tener que utilizar sabiamente combinadas varias herra-

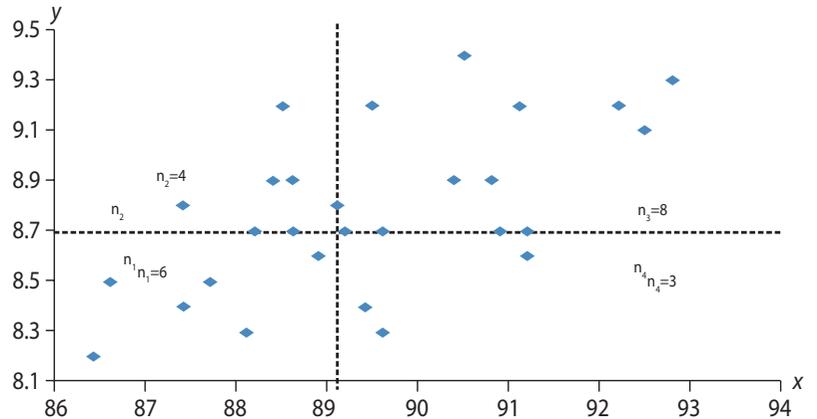


Figura 6.7. Relación entre la proporción de pacientes que esperan más de 45 min (y) y la de consultas a demanda (x) (según datos del cuadro 3)

Cuadro 6.4								
Cuadro de la prueba de los signos								
n	0.01	0.05	n	0.01	0.05	n	0.01	0.05
20	3	5						
21	4	5	46	13	15	71	24	26
22	4	5	47	14	16	72	24	27
23	4	6	48	14	16	73	25	27
24	5	6	49	15	17	74	25	28
25	5	7	50	15	17	75	25	28
26	6	7	51	15	18	76	26	28
27	6	7	52	16	18	77	26	29
28	6	8	53	16	18	78	27	29
29	7	8	54	17	19	79	27	30
30	7	9	55	17	19	80	28	30
31	7	9	56	17	20	81	28	31
32	8	9	57	18	20	82	28	31
33	8	10	58	18	21	83	29	32
34	9	10	59	19	21	84	29	32
35	9	11	60	19	21	85	30	32
36	9	11	61	20	22	86	30	33
37	10	12	62	20	22	87	31	33
38	10	12	63	20	23	88	31	34
39	11	12	64	21	23	89	31	34
40	11	13	65	21	24	90	32	35
41	11	13	66	22	24	100	36	39
42	12	14	67	22	25			
43	12	14	68	22	25			
44	13	15	69	23	25			
45	13	15	70	23	26			

La estratificación y el diagrama de dispersión forman parte de las siete herramientas consideradas por algunos autores como las básicas para el control de la calidad y de conocimiento indispensable por todo el personal de una organización con programa de gestión de la calidad.

mientas de una manera secuencial, razonada y con sentido común. De todas las herramientas posibles, hay una cierta coincidencia en destacar siete como las más útiles: las Siete Herramientas; llamadas así por los japoneses, en memoria de las famosas siete armas de Benkei, un guerrero-sacerdote japonés, con las que venció todas sus batallas (en Estados Unidos de Norteamérica algunos autores las han rebautizado como los “7 magníficos”). Estas Siete Herramientas son las que figuran en el cuadro 6.5; entre ellas están, junto al listado de criterios, tanto la estratificación como los diagramas de dispersión, lo cual da una idea de su relevancia potencial. Estas herramientas son consideradas por Ishikawa como correspondientes al nivel elemental, dentro de los tres niveles de métodos estadísticos para el control de la calidad, definidos en función de su dificultad. Son ade-

Cuadro 6.5
Las siete herramientas elementales para el control de la calidad

1. Diagrama de Pareto
2. Diagrama causa-efecto
3. Estratificación
4. Hoja de comprobación de presencia de defectos (listado de criterios para su cuantificación)
5. Histograma
6. Diagrama de dispersión
7. Gráficos de control

Fuente: Adaptado de Ishikawa K.

más, según Ishikawa, indispensables para los programas de calidad y deben de poder ser utilizadas por todo el personal de la organización.

BIBLIOGRAFÍA

- Goldfield N, Pine M, Pine J. Techniques for analyzing processes and making team decisions. En : Measuring and managing health care quality. Gaithersburg: Aspen Pub, 1995. Vol. I, Cap. 5.
- Ishikawa K. EL enfoque estadístico y algunas herramientas estadísticas sencillas. En: Introducción al control de calidad. Madrid: Díaz de Santos, 1994. Cap. 2.
- Rosner B. Fundamentals of Biostatistics. 4ª ed. Belmont: Duxbury Press, 1995.
- Scholtes PR, Joiner BL, Streibel BJ. Getting started: Learning the tools. En: The team handbook. 2ª ed. Madison: Joiner Associates, 1996. Cap. 2.

Anexo 6.1**Indicaciones para la aplicación práctica de los contenidos de este capítulo**

Nota: estas indicaciones complementan las del capítulo 5, en caso de haber elegido como problema por mejorar uno que necesite explorar la importancia de diversas causas hipotéticas.

Indicaciones para la aplicación práctica

Puede escogerse entre medir (cuantificar) directamente las causas hipotéticas a indagar o realizar una cuantificación subjetiva mediante una encuesta de causas. A continuación se dan las indicaciones para ambos casos, aunque normalmente se elige uno de los dos.

I. Cuantificación directa de las causas hipotéticas

- I.1. Definir el listado de causas a verificar. Siguiendo el esquema de componentes del estudio que se diseñó en el capítulo 5 para el criterio o indicador que mide el problema de calidad a mejorar, completando el apartado correspondiente a “criterios a utilizar” con el listado de “causas a verificar”.

Este listado, extraído de las encontradas en el análisis del problema (capítulo 3), contiene las causas que van a ser medidas en el estudio.

- I.2. Revisar los componentes del estudio, agrupando para cada causa el tipo de variable de que se trata (cuantitativa, cualitativa) y de unidad de estudio. Plasmarlo en el formato adjunto.
- I.3. Recogida de datos: al igual que se indica en el capítulo 5, pero midiendo en cada caso el criterio o indicador de calidad que representa el problema a mejorar y las causas hipotéticas que contribuyen a su presencia.

II. Cuantificación subjetiva de las causas hipotéticas

- II.1. Definir listado de causas a realizar. Extraer, de las causas encontradas al analizar el problema, aquellas que queremos valorar.
- II.2. Diseñar una encuesta en la que cada pregunta es la valoración de una de las causas a estudio. Puede utilizarse el esquema de respuesta más simple (tres opciones): “muy importante”, “poco importante”, “nada importante”) o cualquier otro que permita priorizar fácilmente sin ambigüedades.
- II.3. Recogida de datos: administrar la encuesta de causas al personal implicado en el problema a valorar y mejorar.

Nota: la encuesta complementa (no sustituye) el diseño de estudio (capítulo 5) para la medición basal del criterio o indicador que representa el problema de calidad por evaluar. La encuesta de causas tiene como objetivo identificar aquellas sobre las que hay que intervenir, según los implicados en el problema o proceso por mejorar.

Capítulo 7

Análisis y presentación de los datos de una evaluación

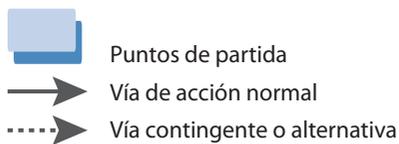
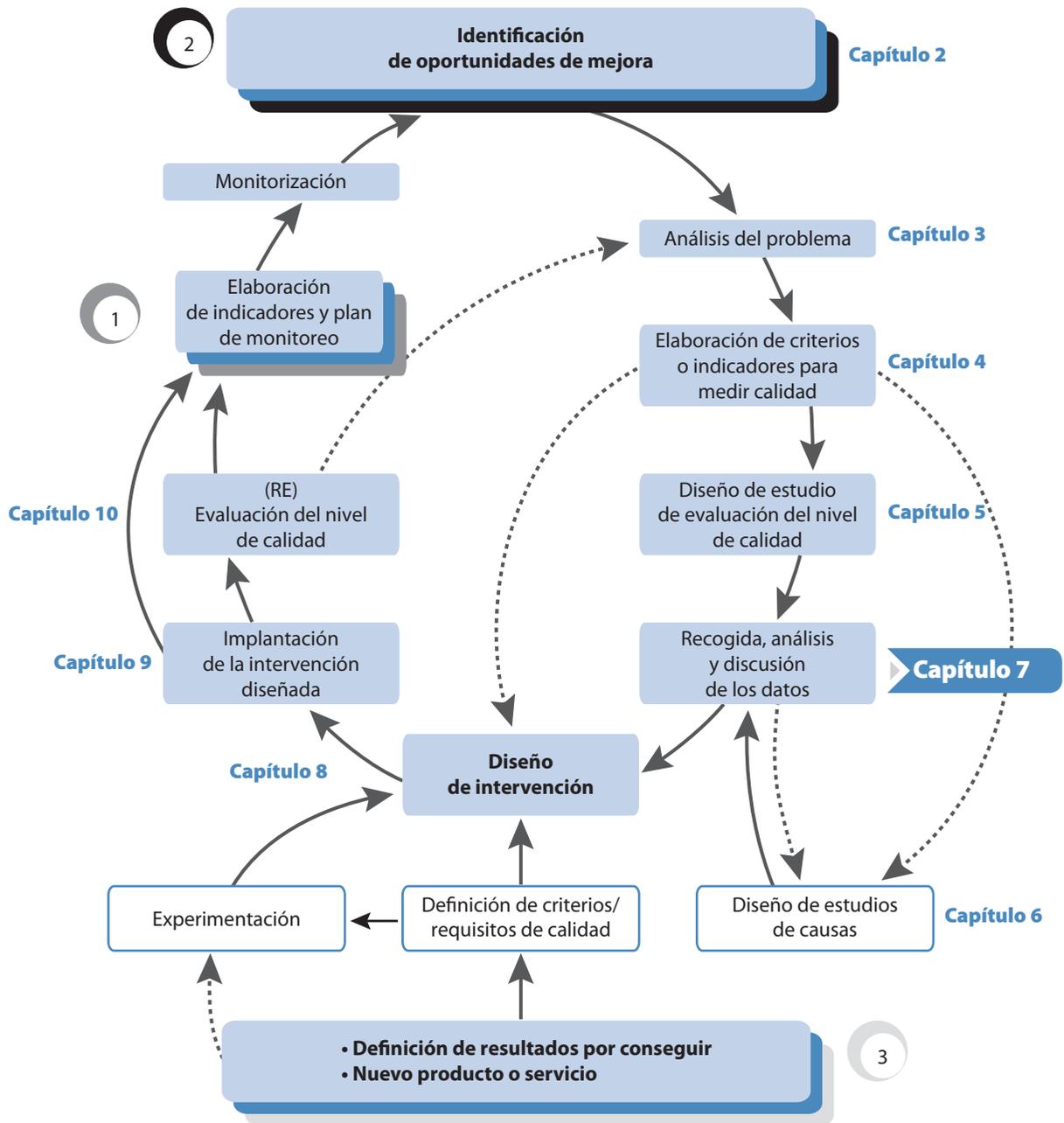
Contenido general y objetivos específicos

1. Introducción
2. Los datos de la evaluación: estimación del nivel de cumplimiento de los criterios y cálculo de intervalos de confianza
3. Influencia del método de muestreo en la estimación de intervalos de confianza
4. Presentación gráfica de los datos de una evaluación
5. Análisis gráficos de los datos para priorizar intervenciones: diagrama de Pareto de los defectos de calidad (incumplimientos de criterios) encontrados
6. Análisis de los datos de estudios de causas hipotéticas

Anexo 7.1. Intervalo (i) para un grado de confianza de 95%, según estimación puntual y tamaño muestral

Anexo 7.2. Indicaciones para la aplicación práctica de los contenidos de este capítulo

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Contenido general

Este capítulo detalla la forma en que los datos obtenidos en el estudio del nivel de calidad deben ser presentados y analizados desde un punto de vista estadístico, y para que sirvan como base para diseñar la intervención que haga mejorar la calidad. Se subraya la necesidad de calcular los intervalos de confianza de los valores encontrados, si se ha utilizado una muestra, y se especifica paso a paso la construcción de un diagrama de Pareto para el análisis de las deficiencias de calidad a subsanar, o también de la frecuencia de las causas hipotéticas indagadas, si ha sido de este tipo el estudio realizado. Se muestra también la construcción de otros tipos de gráficos (estrella, Box and whiskers) que pueden ser utilizados para la presentación global de los datos.

Objetivos específicos

En este capítulo se describen los métodos para:

1. Resumir de manera informativa los datos de una evaluación de la calidad.
2. Calcular los intervalos de confianza de las estimaciones de cumplimiento encontradas.
3. Distinguir los conceptos de "intervalo" y "confianza" e interpretar sus valores.
4. Ordenar los datos de forma adecuada para construir un diagrama de Pareto sobre las deficiencias de calidad o causas hipotéticas encontradas.
5. Construir un diagrama de Pareto completo con los resultados de la evaluación.
6. Explicar los resultados de la evaluación utilizando el diagrama de Pareto.
7. Presentar gráficamente los resultados de la evaluación utilizando un gráfico tipo estrella o radar.
8. Presentar gráficamente los resultados de la evaluación utilizando un gráfico "box and whiskers".
9. Describir correctamente las presentaciones tabular y gráfica de los resultados de una evaluación

Análisis y presentación de los datos de una evaluación

“Evaluar para mejorar no para juzgar”

1. Introducción

Los capítulos 5 y 6 describen la metodología para diseñar estudios orientados a la recogida de datos sobre el nivel de calidad (cumplimiento de criterios o requisitos de calidad, normalmente de atención clínica, capítulo 5) del aspecto o aspectos de los servicios que prestamos que más nos haya interesado y, eventualmente, de las causas hipotéticas de la mala calidad (capítulo 6). Estos datos obtenidos según el plan establecido en los respectivos estudios son la base para decidir qué podemos y debemos hacer para mejorar la calidad en los problemas evaluados. Para ello, es necesario analizar y presentar adecuadamente los datos obtenidos, de forma que sirvan de base para la discusión de las oportunas intervenciones para la mejora. En este capítulo se revisa la forma de realizar en la práctica y con rigor científico esta presentación y análisis de los datos. Los métodos a utilizar incluyen el tratamiento de los datos para la inferencia estadística de los parámetros (esencialmente el nivel de cumplimiento real de los criterios) que hemos medido mediante muestra, y la representación gráfica de los mismos, destacando los niveles de cumplimiento (gráfico de barras y tipo estrella), la variabilidad y precisión de las mediciones (gráfico “box and whiskers”), o bien los incumplimientos o causas hipotéticas encontradas (según el tipo de estudio realizado) para priorizar los aspectos más importantes de la intervención (gráfico de Pareto).

El capítulo se completa con las indicaciones para la aplicación práctica de los métodos estadísticos y gráficos más relevantes para la

presentación y análisis de los datos obtenidos en los estudios diseñados según la metodología descrita en los capítulos 5 y 6.

2. Los datos de la evaluación: estimación del nivel de cumplimiento de los criterios y cálculo de intervalos de confianza

La mayoría de las evaluaciones del nivel de calidad se realizan utilizando muestras. En estas condiciones, lo que realizamos no son mediciones directas y sin error posible sino *estimaciones* de los valores reales que tiene la variable (criterio de calidad) que queremos evaluar. Estos valores reales, sean proporciones o medias, se conocen como *parámetros* en la jerga estadística. Dicho de una forma más precisa, lo que hacemos (o debemos hacer) cuando medimos el nivel de calidad con una muestra es realizar una *inferencia estadística* sobre las características (parámetros) de poblaciones o grupos amplios de casos, sobre la base de información aportada por la muestra; siempre que, recordémoslo, la muestra sea aleatoria (representativa), no sesgada, como veíamos en el capítulo 5 al hablar del diseño del estudio. Las estimaciones a inferir sobre el parámetro que nos interese pueden ser de tipo puntual (un solo valor) o de intervalos de valores.

2.1. Estimación puntual del nivel de cumplimiento de un criterio

La primera *inferencia* que podemos hacer con los datos de la muestra es la propia estimación del nivel de cumplimiento que resulta en la medición realizada para cada uno de los criterios.

Los datos de una muestra representativa nos sirven para estimar por medio de inferencia estadística el parámetro que medimos.

La primera inferencia a realizar es la estimación puntual del parámetro (por ejemplo, proporción de cumplimiento de un criterio) que hemos medido. Para ello hacemos los cálculos con los datos de la muestra como si se tratase de la totalidad del universo de donde se ha extraído la muestra.

La estimación puntual puede y debe completarse con la estimación del intervalo de confianza; un intervalo de valores dentro del cual tengamos un cierto nivel de seguridad (confianza) de que se encuentre el valor real.

El intervalo de confianza implica dos aspectos que, aunque en la práctica han de ir necesariamente juntos, conviene entender y considerar por separado; el intervalo (conjunto de valores entre los cuales estimamos que está el valor real) y la confianza (probabilidad de que sea cierto que el intervalo contenga al valor real).

El intervalo de confianza (estimación puntual $\pm i$) se calcula fácilmente a partir de los datos de la muestra, una vez fijada la confianza y estimando el error llamado error estándar ($i = z \cdot \text{error estándar}$, siendo z un valor dependiente de la confianza).

Por ejemplo, si hemos evaluado la proporción de hipertensos correctamente diagnosticados en una muestra aleatoria de 60 casos y hemos encontrado que 45 de estos cumplen el criterio de diagnóstico correcto, podemos *inferir o estimar* que una proporción de $45/60 = 0.75$ o un porcentaje del 75% de los hipertensos están correctamente diagnosticados. Esta estimación se conoce como *estimación puntual* del parámetro (valor poblacional real) que queremos medir, en este caso la proporción de hipertensos correctamente diagnosticados.

Sin embargo, todos sabemos que al medir algo por medio de una muestra el resultado está sujeto a errores propios del muestreo y que, por tanto, la estimación puntual puede no ser exactamente igual al valor real que encontraríamos si midiésemos no una muestra sino todos los casos. Es más, no sólo somos conscientes de que la estimación puntual puede no ser igual al valor real, sino que también está claro que no sabemos, dando sólo la estimación puntual, lo cerca o lo lejos que puede encontrarse del valor real. Para solucionar este problema hemos de completar nuestro proceder de inferencia estadística estimando no sólo un valor (estimación puntual) sino un intervalo de valores, dentro de los cuales podamos tener una cierta seguridad de que se encuentra el valor real, poblacional, de lo que hemos medido a través de una muestra.

2.2. Estimación del intervalo de confianza del cumplimiento de un criterio

Esta segunda estimación es lo que se conoce como *intervalo de confianza* de la medición realizada. El intervalo de confianza expresa la precisión o exactitud probable (¡todo en estadística es probabilidad!) de la medición realizada, en relación con el valor real; obviamente, cuanto más amplio sea el intervalo, menos precisa es nuestra estimación. En la amplitud del intervalo intervienen, como luego veremos, el tamaño de la muestra, la variabilidad de los datos de la muestra y la confianza (certeza o probabilidad) que queramos tener de que sea cierto que este intervalo contiene el valor real.

No debe de extrañarnos la afirmación de que las estimaciones de cumplimiento de criterios en base a muestras deberían incluir *siempre* la estimación del intervalo de confianza. Pero, ¿cómo podemos estimar este intervalo de valores?; es fácil: sólo necesitamos los datos de nuestro muestreo y decidir sobre la *confianza* que queremos tener en nuestra estimación. De hecho, conviene entender que, aunque relacionados, el concepto *intervalo de confianza* entraña dos aspectos diferentes: el *intervalo* y la *confianza*. El intervalo es la gama de valores dentro de la cual estimamos que está el valor real (poblacional) del parámetro evaluado (en nuestro caso este parámetro sería la proporción de cumplimientos de un criterio), mientras que la confianza es la probabilidad de que sea *cierto* que el intervalo contiene el valor real del parámetro evaluado. Sin embargo, dado que ambos aspectos van necesariamente unidos en la inferencia estadística (para una misma muestra, si aumentamos la confianza de la estimación aumenta la amplitud del intervalo), es costumbre considerarlos como un concepto único. Vamos a ver cómo se calcula.

2.3. Cálculo del intervalo de confianza del nivel de cumplimiento de un criterio

El razonamiento y fórmula general para estimar el intervalo de confianza del parámetro medido con muestra es: estimación puntual $\pm i$, siendo $i = z \cdot \text{Desviación estándar del parámetro estimado}$

donde z es un valor de la distribución normal (la distribución de probabilidades más conocida en estadística) relacionado con la *confianza* que queramos tener en la estimación (probabilidad de que el intervalo contenga el valor real del parámetro estimado); y la *desviación estándar del parámetro estimado* es lo que se conoce también como error del muestreo o error estándar, que se estima con base en la desviación estándar de los datos de la muestra (una expresión de su dispersión o variabilidad) y el propio tamaño de la muestra. Así pues, el cálculo del intervalo de confianza entraña los

siguientes dos pasos previos: (1) determinar el valor de z y (2) calcular el error estándar.

1. Determinar el valor de z . Para ello hay que decidir sobre la confianza que queremos tener para la estimación; es decir, lo seguros que queramos estar de que el valor real esté contenido en el intervalo que vamos a estimar. Está claro que una confianza del 100% exigiría medir todos los casos, no una muestra, o bien dar como intervalo toda la gama posible de valores, lo cual no tendría ninguna utilidad como estimación. En consecuencia, hay que *arriesgarse* algo; habrá que aceptar una cierta probabilidad de equivocarse; es decir, estimar un intervalo de valores con una confianza inferior al 100%. ¿Cuál debe ser esta confianza? En general, se considera una confianza aceptable si ésta es del 95%. Es decir, si tenemos un 95% de probabilidades de que el intervalo contenga el valor real, y sólo un 5% de probabilidades de que el valor real sea un valor externo al intervalo (probabilidad de equivocarnos en nuestra estimación).

Para una confianza de 95%, z es 1.96. Este valor es el que aparece en las tablas de la distribución normal, para esta confianza. El cuadro 7.1 contiene los valores de z , correspondientes a los diversos niveles de confianza más frecuentemente empleados.

Nivel de confianza*	Z
90%	1.64
95%	1.96
99%	2.58

2. Calcular el error estándar o desviación estándar del parámetro estimado. Anotado habitualmente como s_p (cuando el parámetro estimado es una proporción, como ocurre en la mayoría de los criterios de calidad), o como $s_{\bar{x}}$ (cuando el parámetro estimado es una media). En el caso de la estimación del nivel de calidad como proporción de cumplimiento de un criterio, la fórmula de cálculo es:

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

donde p es la estimación puntual de la proporción de cumplimiento, y n es el tamaño de la muestra utilizada.

En el ejemplo empleado de la estimación de la proporción de hipertensos correctamente diagnosticados, teníamos una estimación puntual del 75% en una muestra de 60; la estimación del intervalo de confianza para el cumplimiento de este criterio sería, con una confianza de 95%:

$z = 1.96$ (como corresponde a una confianza de 95%)

$$p = 0.75 \quad 1-p = 0.25 \quad n = 60$$

$$s_p = \sqrt{\frac{(0.75)(0.25)}{60}}$$

con lo cual

$$i = z \cdot s_p = 1.96 \sqrt{\frac{(0.75)(0.25)}{60}} = 0.11$$

y el intervalo de confianza: *estimación puntual* $\pm i = 0.75 \pm 0.11$

¿Qué quiere decir esto?, pues que estamos un 95% seguros de que el valor real de cumplimiento del criterio es un valor que está entre 0.64 (límite inferior del intervalo: $0.75 - 0.11$) y 0.86 (límite superior del intervalo: $0.75 + 0.11$); en porcentaje: entre 64% y 86%.

Por otra parte, sólo con ver la fórmula para su cálculo (n está en el denominador de s_p) es evidente que cuanto mayor sea el tamaño de la muestra, más pequeño va a resultar el intervalo; es decir: más precisa puede ser nuestra estimación (ver anexo 7.1). De esta forma, si al estimar el nivel de cumplimiento de un determinado criterio el intervalo de confianza nos resulta demasiado amplio para decidir sobre su mejora, podemos obtener una mayor precisión ampliando la muestra; siempre que los

Es habitual estimar el intervalo de confianza para una confianza de 95%, con lo cual en la fórmula para su cálculo $z=1.96$.

Una vez calculada la desviación estándar del parámetro estimado (error estándar), se puede estimar el intervalo de confianza. Este intervalo nos indica los valores mínimo y máximo entre los cuales se encuentra el valor real del parámetro estimado, con una confianza (seguridad de que sea cierto) prefijada que habitualmente decidimos que sea del 95%.

Para una misma confianza, si aumentamos el tamaño de la muestra, más precisión obtendremos al estimar el intervalo; es decir más pequeña es la gama de valores entre los cuales estimamos que está el valor real.

casos adicionales los obtengamos también de forma aleatoria y del mismo marco muestral. El anexo 7.1 contiene los cálculos de i para las diversas estimaciones puntuales de porcentajes y determinados tamaños de muestra. La influencia del tamaño del universo o marco muestral es, la mayoría de las veces, irrelevante, tal como veremos a continuación.

2.4. Influencia del tamaño del universo o marco muestral

¿Qué ocurre si extraemos la muestra de marcos muestrales (total de casos, universo o población) de tamaños diferentes?, ¿influye el tamaño del marco muestral en las estimaciones que podamos inferir? La respuesta es que esto no siempre sucede.

En principio, el tamaño del universo muestral no influye a efectos prácticos a menos de que la muestra sea de un tamaño superior al 10% del total del universo o marco muestral de donde se ha extraído. En este caso, si bien no hay influencia en cuanto a la estimación puntual, sí que conviene recalcular la estimación del intervalo de confianza porque, manteniendo la misma confianza, el intervalo va a resultar más pequeño: es decir, la estimación va a ser más precisa.

El cálculo del intervalo ajustado por el tamaño del universo se realiza multiplicando por la fracción:

$$\sqrt{\frac{N - n}{N - 1}}$$

donde N es el tamaño del universo y n es el tamaño de la muestra.

La fórmula es entonces:

$$i' = i \cdot \sqrt{\frac{N - n}{N - 1}}$$

donde i es el cálculo realizado sin tener en cuenta el tamaño del universo, tal como hemos visto antes. En nuestro ejemplo de estimación de la proporción de hipertensos correctamente diagnosticados, aplicada en una muestra de 60 casos, realizaremos este ajuste para la estimación del intervalo de confianza siempre que el tamaño del universo (total de hipertensos)

sea inferior a 600, puesto que en esos casos el tamaño de la muestra (60) va a ser mayor que el 10% del tamaño del universo.

Supongamos, por ejemplo, que el total de hipertensos diagnosticados en nuestro centro (N , marco muestral o universo para nuestro estudio) es de 300; el intervalo de confianza, ajustado por el hecho de que $n > 10\% N$, sería

$$i' = i \cdot \sqrt{\frac{300 - 60}{299}} = 0.10$$

lo cual quiere decir que podemos afirmar, con una confianza del 95%, que la proporción real de cumplimiento es un valor comprendido entre 65% y 85% (0.75 ± 0.10), en lugar de 64% y 86% como teníamos antes.

Obsérvese, que para $n=N$ (todos los casos), la fracción

$$\sqrt{\frac{N - n}{N - 1}}$$

es cero, lo cual es lógico porque no hablaríamos entonces de inferencia estadística sino de medición directa del parámetro; de igual manera, a medida que la diferencia entre N y n se agranda, la fracción de ajuste tiende a ser 1, lo cual indica que el tamaño del universo no va a influir en la estimación del intervalo de confianza. De hecho, como regla, no merece la pena hacer el cálculo ajustado para cualquier $n < 10\% N$.

2.5. Inferencia estadística sobre medias

La mayoría de los criterios para medir nivel de calidad son cualitativos categóricos: se mide en cada caso la cualidad de si lo cumple o no; clasificamos los casos en estas dos categorías (los que cumplen y los que no), y estimamos la proporción de una de ellas, la que más nos interese. Sin embargo, en ocasiones el nivel de calidad se mide con criterios que son variables cuantitativas cuyo parámetro característico es una media (por ejemplo media de minutos por consulta, media de minutos que tiene que esperar el paciente antes de ser atendido, etc.).

En estos casos, aunque el razonamiento estadístico es igual, las fórmulas para la inferencia de las estimaciones tanto puntuales como de intervalo son diferentes. El cuadro 7.2 contiene las fórmulas necesarias, tanto para el caso

El tamaño del universo (N) de donde se extrae la muestra (n) no influye en la estimación del intervalo de confianza a menos que $n > 10\% N$.

Si el criterio mide una cantidad cuyo parámetro resumen es una media, las fórmulas para la estimación puntual y del intervalo de confianza son peculiares, aunque el razonamiento estadístico es idéntico al visto para proporciones.

Cuadro 7.2
Inferencia estadística en base a muestras no sesgadas:
Estimaciones puntuales e intervalos de confianza de proporciones y medias

Parámetro a estimar	Tamaño de muestra (n)	Estimación puntual	Intervalo de confianza (95%) (estimación puntual ± i)*
- Proporción (p) ejemplo: proporción de cumplimientos de un criterio de calidad.	Menor que el 10% del tamaño del marco muestral (N)	$p = \frac{\text{n}^\circ \text{ cumplimientos}}{n}$	$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$
	Mayor que el 10% del tamaño del marco muestral (N)	$p = \frac{\text{n}^\circ \text{ cumplimientos}}{n}$	$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
- Media (\bar{x}) (ejemplo: media de minutos por consulta)	Menor que el 10% del tamaño del marco muestral (N)	$\bar{x} = \frac{\sum X_i}{n}$	$\bar{x} \pm 1.96 \frac{S}{\sqrt{n}}$
	Mayor que el 10% del tamaño del marco muestral (N)	$\bar{x} = \frac{\sum X_i}{n}$	$\bar{x} \pm 1.96 \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

X_i : los diversos valores individuales de cada caso incluido en la muestra

s: desviación estándar de los datos de la muestra = $\sqrt{\frac{\sum (X_i - \bar{x})^2}{n-1}}$

* La fórmula general para calcular los intervalos de confianza es $i = z \cdot \text{Desviación estándar del parámetro estimado}$, donde z es el valor en la distribución normal que corresponde a los diversos niveles de confianza (ver cuadro 7.1), y la desviación estándar del parámetro corresponde al llamado *error estándar*. Si el tamaño de la muestra no es suficientemente grande ($n > 30$), la distribución normal no puede asumirse y se debe emplear el valor de t (de la distribución de la t de student) en lugar de z; t varía con el tamaño de la muestra.

de proporciones (como en nuestro ejemplo de hipertensos bien diagnosticados) como de medias y puede utilizarse como referencia para los casos en que vaya a ser necesario utilizar medias en vez de proporciones.

3. Influencia del método de muestreo en la estimación de intervalos de confianza

En el capítulo 5 vimos que el muestreo puede realizarse con diversos métodos, de los cuales destacamos por su mayor aplicabilidad práctica en general al aleatorio simple, el sistemático y el estratificado (proporcional y no proporcional). El emplear uno u otro método tiene influencia en la variabilidad de los datos de la muestra, y por tanto en las estimaciones que hagamos con ellos, sobre todo en la estimación del intervalo de confianza. Esta influencia puede ser particularmente importante en el muestreo estratificado, y se basa en el hecho de que

al estratificar controlamos uno de los factores (aquel por el que estratificamos) con influencia potencial en la variabilidad de los datos (y por tanto en la desviación estándar). La variabilidad dentro de cada estrato es de esperarse que sea menor (es más homogéneo) que la que exista entre los estratos y por tanto en la muestra que podríamos realizar sin estratificar; por otra parte, las diferencias de tamaño de cada estrato también han de ser tenidas en cuenta, sobre todo en el muestreo estratificado no proporcional. El resultado es generalmente que en las muestras estratificadas la estimación es más precisa si aplicamos las fórmulas adecuadas que si no tenemos en cuenta la estratificación.

El cuadro 7.3 contiene las fórmulas para las estimaciones puntuales e intervalos de confianza según los diversos métodos de muestreo. La primera impresión al ver las fórmulas para las estimaciones en muestreos estratificados es que son de gran complejidad; cuando en rea-

Cuadro 7.3
Estimaciones puntuales de proporciones (p) y medias (x) e intervalos de confianza según tipo de muestreo utilizado

Tipo de muestreo	Estimación puntual	Intervalo de confianza (95%)
- Aleatorio simple	Ver cuadro 7.2	Ver cuadro 7.2
- Sistemático	Ver cuadro 7.2	Ver cuadro 7.2*
- Estratificado		
Proporcional	Ver cuadro 7.2	$P + 1.96 \sqrt{\frac{1-f}{n} \sum W_h \frac{n_h}{n_h-1} P_h(1-P_h)}$ $\bar{X} + 1.96 \sqrt{\frac{1-f}{n^2} \cdot \sum W_h \cdot s_h^2}$
No proporcional	$P = \frac{1}{N} \sum N_h \cdot P_h$ $\bar{X} = \frac{1}{N} \sum N_h \cdot \bar{X}_h$	$P + 1.96 \sqrt{\sum W_h^2 (1-f_h) \frac{P_h(1-P_h)}{n_h-1}}$ $\bar{X} + 1.96 \sqrt{\sum W_h^2 (1-f) \frac{S_h^2}{n_h}}$

* Algunos autores proponen y razonan otras fórmulas semejantes a las del muestreo estratificado.

f: fracción de muestreo = $\frac{n}{N}$; n = tamaño total de muestra; N = tamaño del universo o marco muestral.

h: número de estratos; n_h = tamaño de la muestra en cada estrato; N_h = tamaño de cada estrato.

Σ : sumatorio de todos los estratos.

P_h : proporción de cumplimientos de cada estrato; f_h = fracción de muestreo en cada estrato.

W_h : proporción de casos de cada estrato (N_h) en el total del universo muestral = $\frac{N_h}{N}$

s_h : desviación estándar en cada estrato; \bar{X}_h = media en cada estrato.

lidad no es así. Son fácilmente programables en hoja de cálculo, e incluso manualmente si preparamos una tabla debidamente ordenada para el cálculo de cada elemento necesario por separado, con una secuencia lógica. Veamos, por ejemplo, como se procedería para el caso de un muestreo estratificado proporcional, probablemente uno de los más frecuentemente empleados al evaluar calidad.

Supongamos que hemos estratificado una muestra de 60 casos entre tres profesionales participantes en un estudio del nivel de calidad. Entre los tres tienen en programa un total (N) de 450 hipertensos.

Uno de ellos tiene 90 hipertensos, otro 160 y el tercero 200. Para estratificar de forma proporcional, asignamos a cada profesional un número de casos para la muestra en proporción a su número total de hipertensos; así tendremos:

n = 450, de los cuales
 90 (20%) son del profesional 1 (estrato 1)
 160 (36%) son del profesional 2 (estrato 2)
 200 (44%) son del profesional 3 (estrato 3)

n = 60, de los cuales extraemos
 20% de entre los hipertensos del profesional 1 ($n_1=12$)
 20% de entre los hipertensos del profesional 1 ($n_1=12$)
 20% de entre los hipertensos del profesional 1 ($n_1=12$)

Al evaluar las muestras de cada estrato, hemos encontrado 4 hipertensos bien diagnosticados en la muestra del profesional 1, 17 en la del profesional 2, y 24 en la del profesional 3; un total de 45 sobre el total de 60 casos evaluados. Con ello estimaríamos en un 75% ($45/60 \times 100$) el porcentaje de hipertensos bien diagnosticados.

dos. Al ser la estratificación proporcional, la estimación puntual se hace de igual forma que si fuera una muestra aleatoria o sistemática simple.

Para estimar el intervalo de confianza, podemos proceder sin tener en cuenta que la muestra ha sido estratificada y encontraríamos, tal como vimos más arriba, un intervalo para una confianza de 95% de 0.75 ± 0.11 , ó 0.75 ± 0.10 si ajustamos por el tamaño del universo dado que $60 > 10\%$ 450:

$$0.11 \cdot \sqrt{\frac{450 - 60}{499}}$$

Sin embargo, si tenemos en cuenta la estratificación haríamos los cálculos según la fórmula del cuadro 7.3 y en el orden que se muestra en la siguiente tabla:

ESTRATO	Proporción del estrato en el total (Wh)	Muestra en el estrato (nh)	Proporción de cumplimiento en el estrato (Ph)	1-Ph	$W_h \cdot \frac{n_h}{n_h - 1} \cdot P_h (1 - P_h)$
1	0.20	12	4/12=0.33	0,67	0.048
2	0.36	22	17/22=0.77	0,23	0.067
3	0.44	26	24/26=0.92	0,08	0.034
		60			0.146

$$f = \frac{n}{N} = \frac{60}{450} = 0.13$$

Una vez hecho esto, podemos aplicar la fórmula indicada en el cuadro 7.3 para las muestras estratificadas proporcionales:

$$S_{p,estr.} = \sqrt{\frac{1-f}{n} \sum W_h \frac{n_h}{n_h - 1} P_h (1 - P_h)} = \sqrt{\frac{1 - 0.13}{60} \cdot 0.146} = 0.046$$

y con ello el intervalo, para una confianza del 95% sería:

$$0.75 \pm (1,96 \cdot 0,046) = 0.75 \pm 0.09$$

Una estimación más precisa que la de 0.75 ± 0.11 o 0.75 ± 0.10 que habríamos estimado sin tener en cuenta la estratificación.

En general, el incremento de precisión que comporta la estratificación es mayor cuando la desigualdad entre los estratos sea también mayor, de forma que si las estimaciones pun-

tuales por cada estrato son parecidas, no merece la pena realizar los cálculos del intervalo de confianza teniendo en cuenta la estratificación. Sin embargo, si la estratificación que hemos realizado no es proporcional a los tamaños de los estratos, y encontramos proporciones de cumplimiento diferentes en los diferentes estratos, no es sólo conveniente sino *obligatorio* tenerla en cuenta al menos para la estimación puntual global. Veámoslo con un ejemplo: supongamos decidimos estratificar la muestra de 60 casos para los mismos 3 profesionales del ejemplo anterior extrayendo aleatoriamente 20 casos de entre los hipertensos de cada uno (estratificación no proporcional). En los 20 casos del profesional 1 encontramos 12 correctamente diagnosticados (proporción de 0.6), 15 en los del profesional 2 (proporción de 0.75), y 18 en los del profesional 3 (proporción de 0.9); un total de 45 cumplimientos pero con proporciones diferentes en los estratos. Para hacer nuestra estimación del nivel de cumplimiento de los 3 profesionales conjuntamente, no sería correcto decir 45/60 sino que utilizaríamos la fórmula indicada en el cuadro 7.3 para muestras estratificadas no proporcionales:

$$P = \frac{1}{N} \sum N_h \cdot P_h'$$

cuyos valores en nuestro ejemplo serán los siguientes:

Tamaño de cada estrato (Nh)	Proporción de cumplimiento en cada estrato (Ph)	Nh . Ph
Estrato 1: 90	12/20=0.6	54
Estrato 2: 160	15/20=0.75	120
Estrato 3: 200	18/20=0.90	180
N=450		354

De modo que: $P = \frac{1}{450} \cdot 354 = 0.787$; es decir, 78.7%.

Recuérdese que sin tener en cuenta que la muestra ha sido estratificada no proporcional, la estimación de cumplimiento hubiera sido $45/60 = 0.75$, o 75%.

Hay otros tipos de muestreo no contemplados en este manual (por ejemplo el muestreo en conglomerados y polietápico) que con-

El cálculo del intervalo de confianza para muestras estratificadas no es más difícil que para los otros tipos de muestreo, aunque pueda llevar más tiempo, se puede realizar con calculadora de mano. No obstante, puede obviarse si las estimaciones puntuales en cada estrato son semejantes.

Si se ha realizado un muestreo estratificado no proporcional y las estimaciones puntuales (proporciones de cumplimiento) en cada estrato son diferentes, la estimación puntual global debe realizarse con la fórmula apropiada, teniendo en cuenta los datos de la estratificación.

llevan sus propias fórmulas para una correcta inferencia estadística. En la práctica de los programas de calidad lo normal es que utilicemos en la mayoría de los casos uno u otro de los métodos de muestreo que hemos revisado, pero debe quedar clara la idea de que, en cualquier caso, es conveniente reflexionar para la interpretación de los datos si el tipo concreto de muestreo que hemos utilizado puede o no influir en la forma de realizar los cálculos para la inferencia estadística.

El análisis de los datos de una evaluación se completa con su representación gráfica. Vamos a elegir el tipo de gráfico más adecuado en función del tipo de variable (cumple/no cumple o cantidad de algo) con la que medimos el criterio, el interés que tengamos en representar las estimaciones de los intervalos de confianza, y si queremos representar niveles de cumplimiento o analizar los incumplimientos.

4. Presentación gráfica de los datos de una evaluación

Una correcta estimación del nivel de cumplimiento de los criterios es fundamental para dar base real al conocimiento sobre el nivel de calidad de los aspectos evaluados y evidenciar en qué criterios estamos necesitados de mejora. Sin embargo, en calidad existe siempre una representación gráfica que visualiza y hace más comprensible el resultado de las mediciones.

De entre todas las representaciones gráficas posibles vamos a seleccionar las más útiles y prácticas, de entre las cuales elegiremos la que más nos interese, en función de lo que queramos enfatizar en relación a la evaluación realizada. La selección de uno u otro gráfico está representada en la figura 7.1 en forma de algoritmo decisional; como puede verse, una vez construida la tabla de presentación de los datos, las decisiones que afectan a la elección del gráfico son:

- i) el tipo de variable con que medimos el criterio (categórica, o cuantitativa); (ii) si nos interesa representar sólo estimaciones puntuales o también intervalos; y (iii) si queremos representar cumplimientos o analizar incumplimientos (defectos de calidad). En relación a esta última decisión, las opciones que tenemos para representar cumplimientos son el clásico gráfico de barras o el gráfico tipo estrella o radar (si queremos representar sólo las estimaciones puntuales), y el gráfico “box and whiskers” (si queremos representar los intervalos de confianza); para analizar los

El gráfico de barras es útil para representar las estimaciones puntuales de los niveles de cumplimiento, sobre todo si ordenamos los criterios de mayor a menor cumplimiento.

incumplimientos, objeto de la intervención para mejorar, el diagrama de Pareto es más ilustrativo y adecuado para evidenciar las prioridades para intervenir, aunque también es posible utilizar estos mismos gráficos.

Vamos a ver en primer lugar los gráficos para representar cumplimientos y nos detendremos después a revisar paso a paso la construcción e interpretación de un gráfico de Pareto sobre defectos o incumplimientos. En todos los casos utilizaremos como ejemplo los datos que figuran en el cuadro 7.4, resultado de una evaluación de los criterios de diagnóstico, anamnesis y exploración física inicial del paciente hipertenso.

4.1. Gráfico de barras de los cumplimientos (figura 7.2)

Un gráfico de barras consiste en la representación alineada en el eje de abscisas de las barras que representan los diversos criterios evaluados, con una altura para cada una de ellas que corresponde al nivel de cumplimiento del criterio que representa, según una escala que se coloca como eje de ordenadas. De esta forma podemos comparar visualmente la diferencia en los niveles de cumplimiento de los distintos criterios. La escala de medida puede ser de números absolutos de cumplimientos (tendrá como valor máximo el total de casos de la muestra) o lo que es más común, porcentajes, yendo entonces de 0 a 100.

La representación de los datos de una evaluación por medio de un gráfico de barras es probablemente la representación gráfica más sencilla y común. Sin embargo, para que sea realmente informativa conviene que ordene los criterios y su correspondiente representación de mayor a menor cumplimiento. En caso contrario, si las representamos por el orden del listado de criterios, va a ser de poca ayuda para comparar más allá de distinguir las que se destaquen como más extremas. La figura 7.2 contiene los gráficos de barras de los datos del cuadro 7.4 con y sin ordenar para que pueda apreciarse la diferencia.

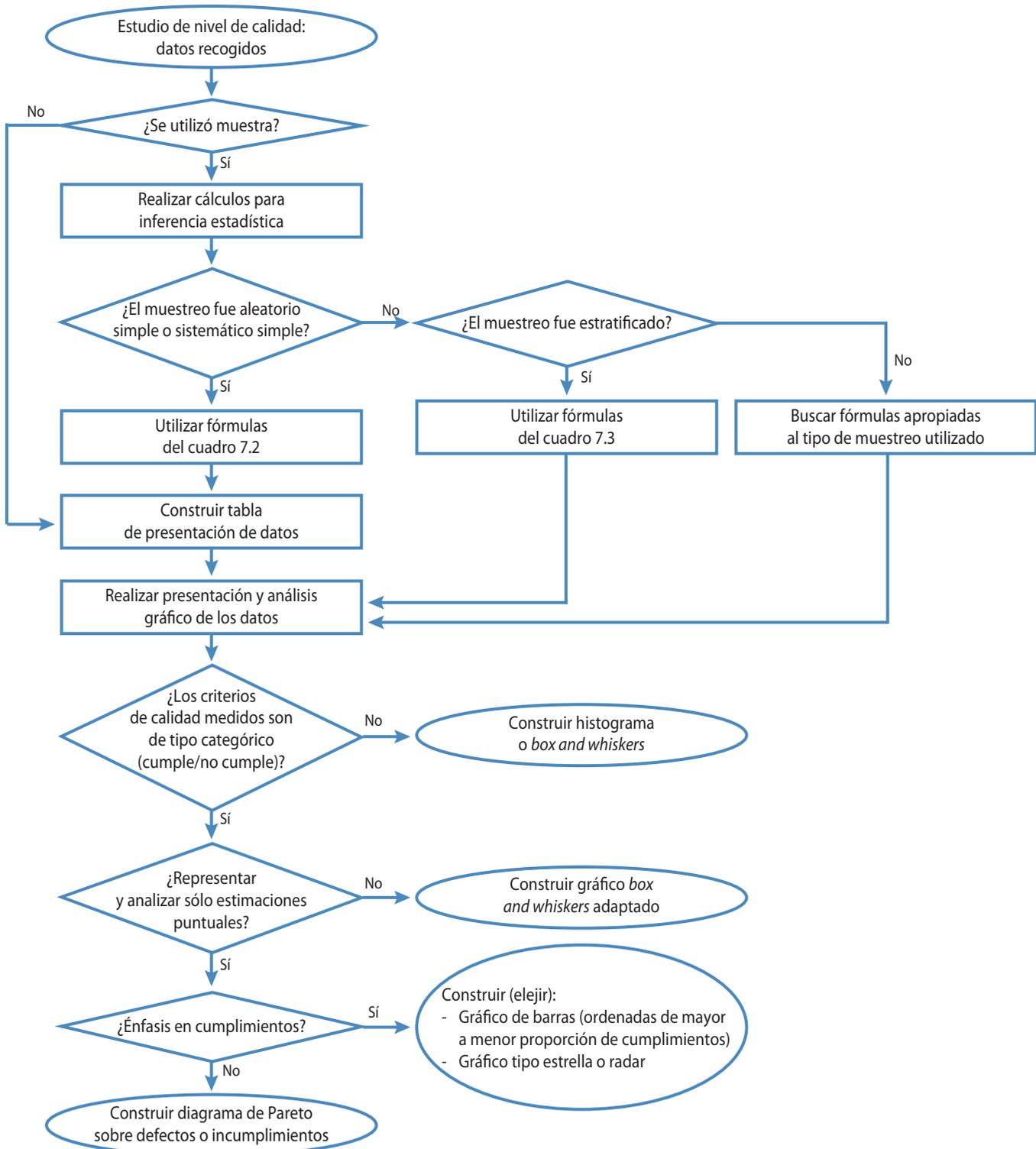


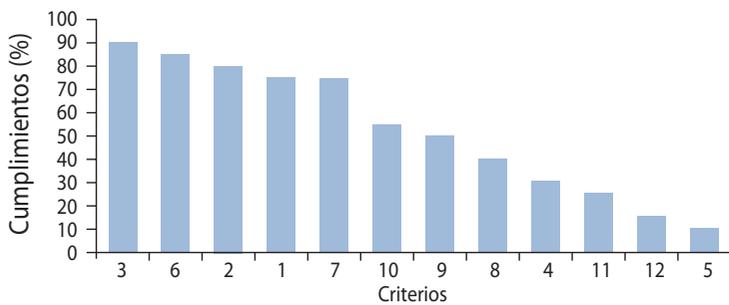
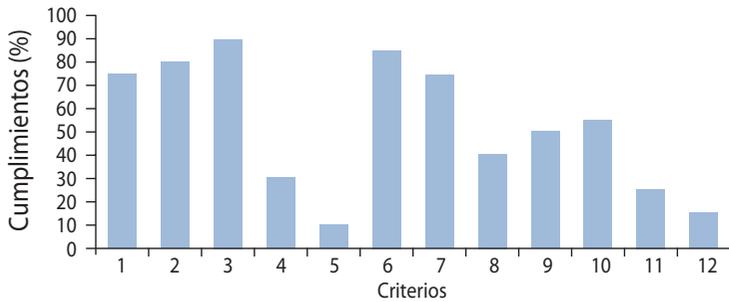
Figura 7.1. Algoritmo decisional para el análisis y presentación de los datos de un estudio del nivel de calidad

Cuadro 7.4

Del diagnóstico, anamnesis y exploración física inicial del paciente hipertenso. Niveles de cumplimiento de los criterios en una muestra aleatoria de 60 casos

Criterio*		% Cumplimiento (Intervalo de confianza del 95%)	
1.	Diagnóstico correcto	75.0	(±11.0)
2.	Antecedentes familiares	80.0	(±10.1)
3.	Antecedentes personales	90.0	(±7.6)
4.	Consumo de tabaco	30.0	(±11.6)
5.	Consumo de alcohol	10.0	(±7.6)
6.	Peso y talla y/o IMC	85.0	(±9.0)
7.	Auscultación cardiaca	75.0	(±11.0)
8.	Exploración abdominal	40.0	(±12.4)
9.	Pulsos periféricos	50.0	(±12.7)
10.	Soplos carotídeos	55.0	(±12.6)
11.	Edemas	25.0	(±11.0)
12.	Fondo de ojo	15.0	(±9.0)

* Este es un enunciado resumido. La descripción detallada de los criterios debe seguir el formato descrito en el capítulo 4.



Criterios

- | | |
|----------------------------|--------------------------|
| 1. Diagnóstico correcto | 7. Auscultación cardiaca |
| 2. Antecedentes familiares | 8. Exploración abdominal |
| 3. Antecedentes personales | 9. Pulsos periféricos |
| 4. Consumo de tabaco | 10. Soplos carotídeos |
| 5. Consumo de alcohol | 11. Edemas |
| 6. Peso y talla y/o IMC | 12. Fondo de ojo |

4.2. Gráfico tipo estrella o radar (figura 7.3)

En este gráfico, cada criterio se representa en un eje cuyo origen es el centro de una estrella o asterisco, de forma que cada eje (cada criterio) es como un brazo del asterisco. La escala de los ejes, todos iguales, puede ser sobre el total de casos de la muestra o como porcentaje (de 0 a 100), que es como hemos representado los datos de nuestro ejemplo. Al representar cada criterio en un eje, puede verse fácilmente la distancia que le queda a cada uno para llegar al 100% de cumplimiento.

Este tipo de gráfico es interesante para comparar evaluaciones de varias áreas, regiones, centros o médicos; cada uno con su "es-

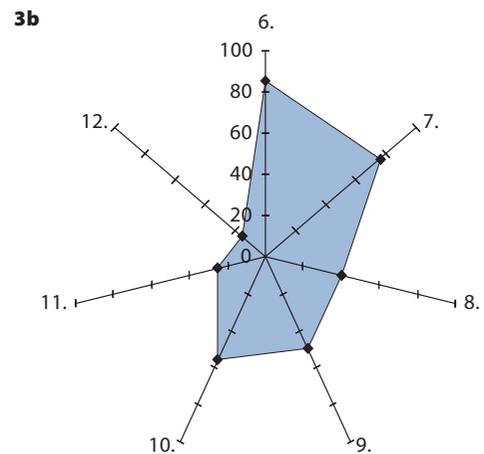
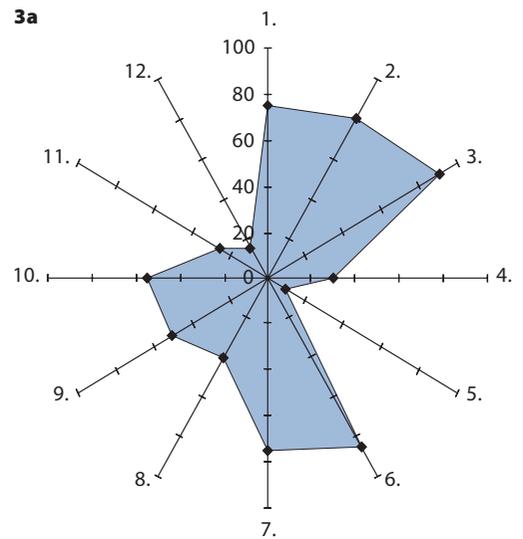


Figura 7.2. Gráfico de barras (datos del cuadro 7.4)

Figura 7.3. Gráfico tipo estrella o radar (datos del cuadro 7.4)

trella” de la forma y tamaño que corresponda a sus resultados, y también para visualizar la mejora sobre un mismo gráfico al comparar dos evaluaciones de una misma área, región o centro, tal como veremos en el capítulo 9. Puede resultar una opción menos atractiva si el número de criterios representados es muy grande o excesivamente pequeño; véase por ejemplo la diferencia entre las figuras 7.3a y 7.3b. En la figura 7.3a se han representado todos los criterios del cuadro 7.4, y en la 7.3b sólo los correspondientes a la exploración física inicial.

4.3. Gráfico *box and whiskers*

La traducción literal sería gráfico de caja y bigotes o gráfico de cajas (*box plot*). Es un tipo de gráfico que se propone a finales de los setenta para visualizar los resultados de una muestra de una variable cuantitativa, como una forma de representación de los llamados por Tukey “cinco números resumen” de los datos recogidos, que son el valor mínimo y el máximo, el primer y tercer cuartil (es decir los valores que corresponden al 25% y 75% de los casos), y la mediana (el valor central de la serie). Los cuartiles delimitan los bordes de la *caja*, en la cual está contenida y representada la mediana, mientras que los valores mínimo y máximo determinan la longitud de los *bigotes*. Con ello se tiene una idea visual de la variabilidad de los datos del criterio o indicador medido, y permite una comparabilidad más clara que si representamos los datos de cada criterio con histogramas (nota: el histograma se vio en el capítulo 3). Este esquema básico es el de la figura 7.4.

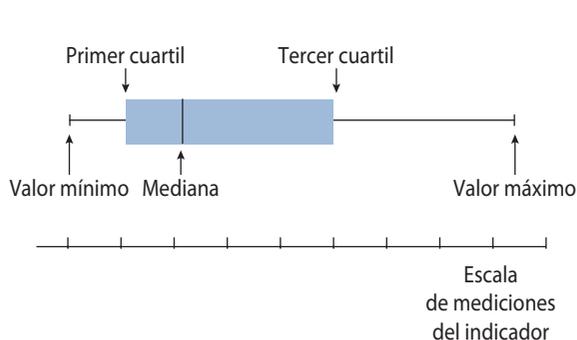


Figura 7.4. Esquema básico del gráfico *box and whiskers*

La idea de representar no solamente las estimaciones puntuales sino también las características que describen la variabilidad de los datos de una muestra ha sido adaptada para múltiples aplicaciones y ha dado lugar a diversas modificaciones de interés. Algunas de ellas son muy sofisticadas, como su utilización para representar las series temporales de las mediciones incorporando en cada medición los resultados de las anteriores, pero otras son sencillas y de una utilidad práctica más inmediata; de estas últimas, las tres más relevantes son probablemente la inclusión de la media en el gráfico, modificar los *bigotes* para evidenciar y representar la existencia de *outliers* y lo que es más importante, para representar las estimaciones de los intervalos de confianza.

Las dos primeras modificaciones son útiles para la representación de criterios cuantitativos (resumidos como media, no como proporción o porcentaje), y consisten en lo siguiente:

Incluir la media. Es simplemente añadir un punto más al gráfico; además de señalar la mediana dentro de la caja, se representa con un punto el valor que corresponde a la media, tal como se ve en la figura 7.5.

Modificar los bigotes para señalar outliers. Si la muestra tiene casos con valores muy extremos, claramente alejados de la mayoría (lo que se conoce como *outliers*) van a producir un gráfico con bigotes muy largos en el lado de los *outliers* y va a proporcionar una información distorsionada de la realidad, al no verse en el gráfico la distribución de valores a lo largo

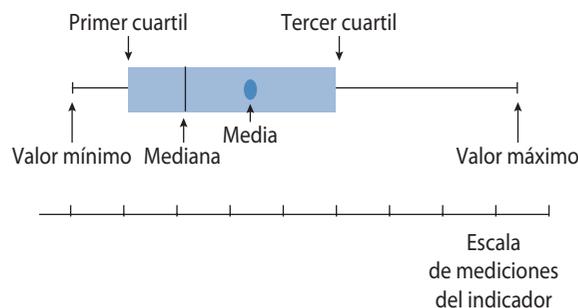


Figura 7.5. Esquema de gráfico *box and whiskers* con mediana y media

El gráfico tipo estrella es útil sobre todo para representar estimaciones puntuales de niveles de cumplimiento y realizar comparaciones. Es menos práctico, más confuso, si el número de criterios representados es grande.

El gráfico *box and whiskers* es útil para representar la variabilidad de los datos de la muestra, además de la estimación puntual del parámetro (proporción o media) de interés. Es el gráfico a elegir si queremos representar los intervalos de confianza y resulta visualmente más informativo si los criterios se ordenan de mayor a menor nivel de cumplimiento.

del bigote. Para evitarlo, se limita la longitud de los bigotes a los valores que comprenden a $1.5 \times$ amplitud *de la caja* (la amplitud de la caja es la distancia entre el primer y tercer cuartil). Los valores que quedan fuera de los bigotes modificados se representan como asteriscos y constituyen los "outliers" de la serie de datos representada. El tipo de gráfico que resulta es el de la figura 7.6.

Representación de los intervalos de confianza. Ésta es la modificación que más nos interesa puesto que es aplicable a la representación del nivel de calidad para cualquier tipo de criterio, sea como media o como proporción de cumplimientos. En esta representación gráfica la caja se delimita en el intervalo de confianza del 68% (aproximadamente la estimación puntual ± 1 desviación estándar del parámetro estimado) y los límites de los bigotes son los valores correspondientes a los límites del intervalo de confianza del 95%. Este esquema es el que se representa en la figura 7.7, y es el que hemos utilizado para representar en la figura 7.8 los datos del cuadro 7.4.

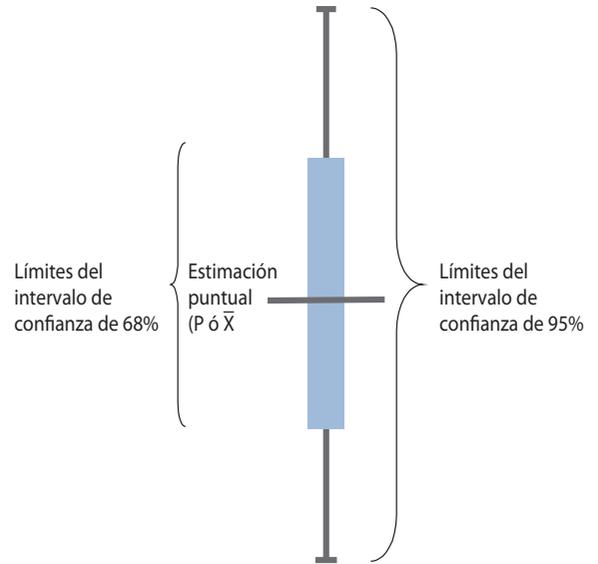


Figura 7.7. Esquema básico para el gráfico *box and whiskers* representando los intervalos de confianza

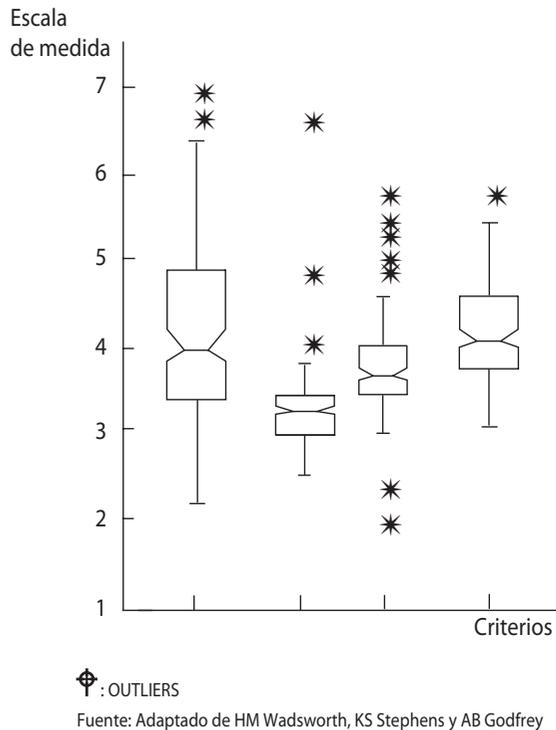
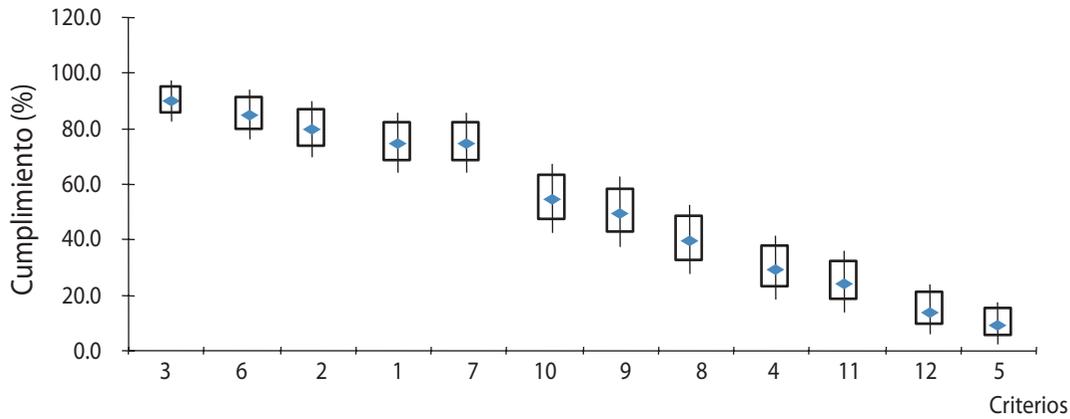


Figura 7.6. Gráfico *box and whiskers* modificado con indicación de outliers

El gráfico *box and whiskers* es una representación estadísticamente más completa que el diagrama de barras, y puede ser mejorada, como el diagrama de barras, ordenando los criterios de mayor a menor cumplimiento, tal como se ha hecho en la figura 7.8. Es posible, igualmente, combinar ambos gráficos superponiendo el de caja y bigotes al diagrama de barras.

De todas formas, aunque todos estos gráficos son útiles para comunicar adecuadamente los resultados de la evaluación, no hay que olvidar que nuestro objetivo al evaluar es saber qué debemos mejorar; en este sentido, nuestra preocupación son fundamentalmente los defectos de calidad o incumplimientos encontrados (las causas del nivel de calidad deficiente), y nos conviene enfocar nuestro análisis gráfico en esta dirección. Para ello, la mejor representación y análisis gráfico de los datos es la realización de un diagrama de Pareto sobre los incumplimientos. Un análisis semejante al que podemos hacer graficando las causas hipotéticas de la mala calidad, tal como se describe en los ejemplos de la encuesta de causas del capítulo 6.



Criterios

- | | |
|----------------------------|--------------------------|
| 1. Diagnostico correcto | 7. Auscultación cardiaca |
| 2. Antecedentes familiares | 8. Exploración abdominal |
| 3. Antecedentes personales | 9. Pulsos periféricos |
| 4. Consumo de tabaco | 10. Soplos carotídeos |
| 5. Consumo de alcohol | 11. Edemas |
| 6. Peso y talla y/o IMC | 12. Fondo de ojo |

Figura 7.8. Gráfico box and whiskers (datos del cuadro 7.4)

5. Análisis gráfico de los datos para priorizar intervenciones: diagrama de Pareto de los defectos de calidad (incumplimientos de criterios) encontrados

El diagrama de Pareto (figura 7.9) es un gráfico de barras que representa la frecuencia de incumplimientos de cada uno de los criterios evaluados, ordenados de mayor a menor número de incumplimientos y completado con una curva de frecuencia porcentual acumulada. El objetivo es visualizar conjuntamente qué criterios aportan mayor número de defectos de calidad (cada incumplimiento es un *defecto de calidad*) y señalar en la curva de frecuencia acumulada cuáles son los que representan la mayoría de defectos y que, por tanto, serían prioritarios como objeto de intervención para mejorar.

Este enfoque de buscar los pocos criterios (o en general causas) que son la base de la mayoría de los defectos de calidad es lo que se co-

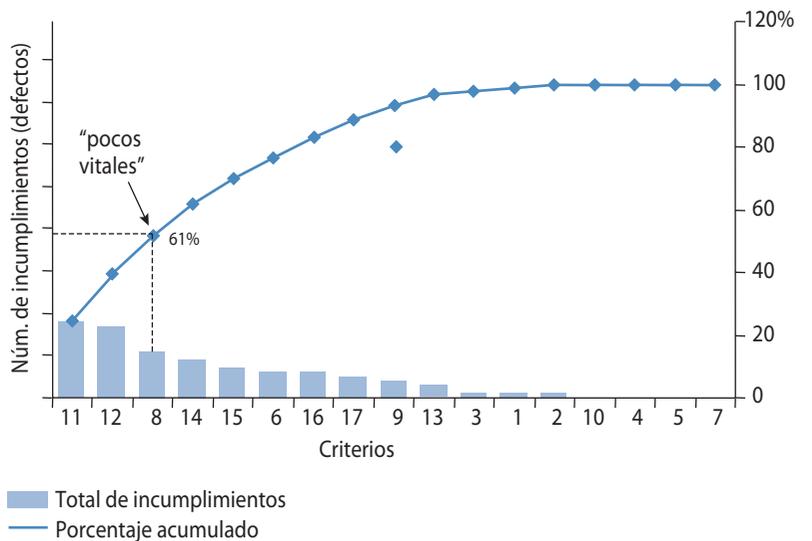


Figura 7.9. Diagrama de Pareto (esquema)

El diagrama de Pareto es un gráfico de barras para representar y analizar las causas de los defectos de calidad (incumplimientos de criterios), aplicando el principio de las “pocas vitales y muchos triviales”, de forma que los criterios se representan ordenadamente de mayor a menor número de incumplimientos y se completa el gráfico con la curva de frecuencia de incumplimientos acumulada para señalar en ella el punto que resume los “pocos vitales”.

Para el diagrama de Pareto, la frecuencia relativa (%) de incumplimientos se calcula sobre el total de incumplimientos encontrados en la muestra, no sobre el total de casos (n) de la muestra.

noce como “principio de Pareto” desde que lo propuso Juran a finales de los años cuarenta.

Juran observó que los estudios de Pareto (un economista italiano de finales del XIX) sobre la distribución de la riqueza señalando su desigualdad y acumulación en unas pocas manos respondían en realidad a una ley universal que se repetía en todos los ámbitos; llevado al terreno de la calidad, indicó que lo más común es que la mayoría de los defectos de calidad se deban a unas pocas causas o tipos de defecto, principio que bautizó como *principio de Pareto* o de *los pocos vitales y muchos triviales*, señalando la importancia de buscar unos pocos defectos o causas de defecto *vitales* que serían la base o acumulación de la mayoría de los defectos de calidad del producto o servicio a mejorar. Para su representación gráfica, aparte de proponer como norma que se ordenaran las barras en el gráfico de mayor a menor importancia en cuanto a número de defectos encontrados y no por cualquier otro orden, propuso completar el análisis con la curva de frecuencia de causas acumuladas (basada en las curvas que propuso Lorenz, también sobre distribución de la riqueza) para señalar en ellas donde se alcanza la mayoría de los defectos.

El principio de Pareto o de los pocos vitales y muchos triviales (ambos en realidad de Juran) se conoce también en la jerga de la calidad en la industria como *principio del 80/20*, indicando que, numéricamente el principio de Pareto quiere decir que el 80% de los defectos se debe a un 20% de las posibles causas.

Si consideramos paso a paso la construcción de un diagrama de Pareto, podemos distinguir dos fases: (i) preparación de los datos; y (ii) construcción del gráfico en sí. Los pasos que implica cada una de las fases se explicitan en el cuadro 7.5. Vamos a realizarlos utilizando los datos de nuestro ejemplo (cuadro 7.4).

5.1. Preparación de los datos

1. Construir una tabla con la frecuencia absoluta (número de incumplimientos) y *relativa* (porcentaje sobre el total de incumplimientos) encontrados en la muestra.

Según los datos del cuadro 7.4, la frecuencia absoluta y relativa de incumplimiento de cada criterio es la que figura en el cuadro 7.6. Estos datos los tendremos normalmente (igual que los de los cumplimientos) como resultado en bruto de la recogida de datos que hayamos hecho. No obstante, también pueden derivarse a partir de las proporciones de cumplimiento como las que contienen el cuadro 7.4: obsérvese que *incumplimientos* es el complemento de *cumplimientos* y que, por ejemplo, un cumplimiento del 90% significa un incumplimiento del 10%; si la muestra es de 60 casos, un 10% de incumplimientos es un total de 6 incumplimientos (un 10% de 60) para ese criterio; de esta forma pueden derivarse criterio a criterio los datos de incumplimiento.

Conviene subrayar que el porcentaje de incumplimientos no se calcula sobre el número de casos de la muestra (como lo haríamos para estimar el porcentaje de cumplimientos de cada criterio) sino sobre el total de incumplimientos (defectos de calidad) encontrados: ésta es la distinción fundamental del diagrama de Pareto sobre los defectos de calidad en relación con el diagrama de barras representando el nivel de cumplimientos. En nuestro ejemplo, cada caso puede contribuir a este total de incumplimientos desde un mínimo de cero (si cumple todos los criterios) hasta un máximo de 12 (si los incumple todos).

2. Ordenar los criterios de mayor a menor número de incumplimientos. Con este orden necesario, el cuadro queda como sigue (cuadro 7.7):

Cuadro 7.5

Construcción de un diagrama de Pareto

A. Preparación de los datos

Hacer una tabla con el listado de criterios y la frecuencia absoluta (nº de veces) y relativa (% sobre el total de incumplimientos de la muestra) de incumplimientos por cada criterio.

Ordenar los criterios de mayor a menor número de incumplimientos.

Calcular la frecuencia relativa (%) acumulada.

B. Construir gráfico

Construir el gráfico de barras, con dos ejes (frecuencia absoluta y relativa) o uno solo (frecuencia relativa).

Construir la curva de frecuencia relativa acumulada.

(Opcional).

Señalar el punto que acumula el 60-80% de los incumplimientos encontrados.

3. Calcular la frecuencia relativa (porcentaje) acumulada.

Este cálculo es la base para representar la curva de Lorenz y poder identificar claramente los pocos *vitales* (aquellos criterios responsables de la mayoría de los defectos encontrados). Implica solamente ir sumando (acumulando) al porcentaje de incumplimientos de cada criterio el porcentaje de incumplimientos que suponen todos los anteriores a él en la tabla. La frecuencia acumulada del último criterio de la lista debe ser, naturalmente, 100. La tabla final de nuestros datos para análisis de Pareto quedaría como se muestra en el cuadro 7.8.

5.2. Construcción del gráfico

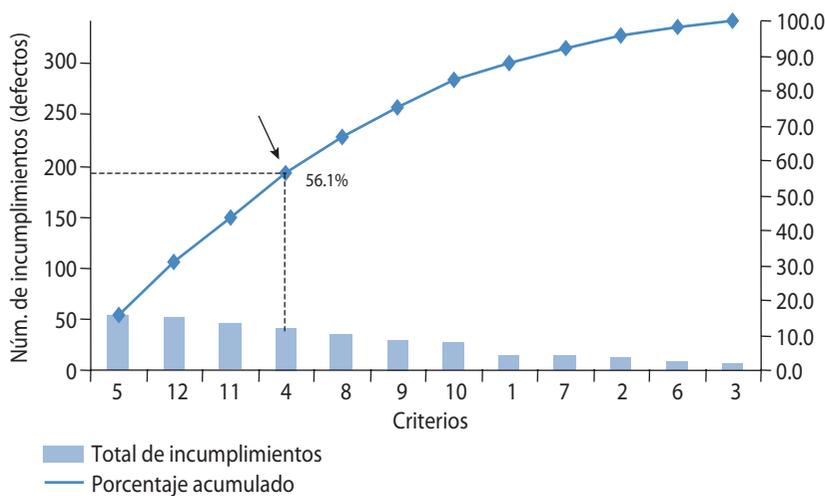
1. Construcción del gráfico de barras con dos ejes (frecuencia absoluta y relativa) o uno solo (frecuencia relativa).

El diagrama de Pareto es un gráfico para comparar tipos de defectos de calidad encontrados y señalar cuáles son los *pocos vitales* para diseñar la intervención para mejorar. Para cumplir este objetivo es suficiente con representar las barras con la frecuencia relativa de incumplimientos de cada criterio (porcentaje sobre total de incumplimiento en la muestra) y la curva de la frecuencia porcentual acumulada, con lo cual sólo necesitaríamos un eje para representar los porcentajes. Sin embargo, no cuesta mucho construir el gráfico más completo, incluyendo la información sobre el número absoluto de defectos que corresponde a cada criterio. Para ello construimos el gráfico con dos ejes: uno a la izquierda para representar la frecuencia absoluta (número de incumplimientos) y otro a la derecha para representar los porcentajes (frecuencia relativa); el eje para la frecuencia absoluta tiene como rango de 0 al total de incumplimientos de la muestra, y el de la frecuencia relativa de 0 a 100. Ambos deben tener el mismo tamaño. Cada barra, representando a cada uno de los criterios, tiene la altura que corresponde, a su vez, al número de incumplimientos (en el eje de la izquierda) y del porcentaje que supone sobre el total (en el eje de la derecha). El eje de la derecha sir-

Cuadro 7.6		
Frecuencia de incumplimientos de los criterios evaluados		
Criterio	Número de incumplimientos (frecuencia absoluta)	Porcentaje (Frecuencia relativa)
1. Diagnóstico correcto	15	4.4
2. Antecedentes familiares	12	3.5
3. Antecedentes personales	6	1.8
4. Consumo de tabaco	42	12.3
5. Consumo de alcohol	54	15.8
6. Peso y talla y/o IMC	9	2.6
7. Auscultación cardiaca	15	4.4
8. Exploración abdominal	36	10.5
9. Pulsos periféricos	30	8.8
10. Soplos carotídeos	27	7.9
11. Edemas	45	13.2
12. Fondo de ojo	51	14.9
	342	100.0

Cuadro 7.7		
Frecuencia de incumplimientos de los criterios evaluados ordenados de mayor a menor		
Criterio	Número de incumplimientos (frecuencia absoluta)	Porcentaje (Frecuencia relativa)
5. Consumo de alcohol	54	15.8
12. Fondo de ojo	51	14.9
11. Edemas	45	13.2
4. Consumo de tabaco	42	12.3
8. Exploración abdominal	36	10.5
9. Pulsos periféricos	30	8.8
10. Soplos carotídeos	27	7.9
1. Diagnóstico correcto	15	4.4
7. Auscultación cardiaca	15	4.4
2. Antecedentes familiares	12	3.5
6. Peso y talla y/o IMC	9	2.6
3. Antecedentes personales	6	1.8
	342	100.0

Cuadro 7.8 Frecuencia absoluta, relativa (ordenadas) y acumulada de los incumplimientos de los criterios evaluados			
Criterio	Número de incumplimientos (frecuencia absoluta)	Porcentaje (Frecuencia relativa)	Frecuencia acumulada
5. Consumo de alcohol	54	15.8	15.8
12. Fondo de ojo	51	14.9	30.7
11. Edemas	45	13.2	43.9
4. Consumo de tabaco	42	12.3	56.1
8. Exploración abdominal	36	10.5	66.7
9. Pulsos periféricos	30	8.8	75.4
10. Soplos carotídeos	27	7.9	83.3
1. Diagnóstico correcto	15	4.4	87.7
7. Auscultación cardiaca	15	4.4	92.1
2. Antecedentes familiares	12	3.5	95.6
6. Peso y talla y/o IMC	9	2.6	98.2
3. Antecedentes personales	6	1.8	100.0
	342	100.0	



Criterios

- | | |
|----------------------------|--------------------------|
| 1. Diagnóstico correcto | 7. Auscultación cardiaca |
| 2. Antecedentes familiares | 8. Exploración abdominal |
| 3. Antecedentes personales | 9. Pulsos periféricos |
| 4. Consumo de tabaco | 10. Soplos carotídeos |
| 5. Consumo de alcohol | 11. Edemas |
| 6. Peso y talla y/o IMC | 12. Fondo de ojo |

Figura 7.10. Diagrama de Pareto (datos del cuadro 7.8)

ve también para referencia de los puntos de la curva de frecuencia acumulada.

2. Completar el diagrama de barras con la curva de frecuencia relativa acumulada.

3. Señalar el punto que acumula la mayoría (en torno a 60% o más) de los incumplimientos.

El diagrama de Pareto que corresponde a nuestros datos es el de la figura 7.10:

Como vemos, hay 4 criterios que representan el 56.1% de los defectos de calidad (incumplimientos) encontrados y serían los *pocos vitales* sobre los que debemos actuar prioritariamente en nuestro esfuerzo para mejorar. En ocasiones este porcentaje de los pocos vitales es mayor, hasta cumplirse o sobrepasar la ley del 80/20.

El diagrama de Pareto es una forma sencilla, brillante e inteligente de presentar y analizar los datos de una evaluación, mostrando dónde está el problema; es decir, lo que tenemos que mejorar. Sin embargo, es conveniente hacer algunas puntualizaciones para su construcción y uso:

5.3. Observaciones sobre la construcción e interpretación de un diagrama de Pareto sobre defectos de calidad

En primer lugar hay que indicar que se asume que todos los criterios y sus correspondientes defectos son igualmente importantes; cualquier valoración sobre la diferente relevancia de uno u otro tipo de defecto de calidad debe hacerse al margen del análisis con gráfico de Pareto, y podrá, obviamente, modificar las prioridades que marca el diagrama.

La segunda observación es de tipo formal: cuando se representan un gran número de criterios o causas hipotéticas, es habitual que varios de ellos tengan un número pequeño de incumplimientos y se agrupen en una sola barra como otros al final del diagrama; en estos casos hay que asegurarse siempre que el número de incumplimientos de cualquier criterio agrupado en otros es menor que el último representado de forma independiente. Por otra parte, no hace buen efecto una barra de otros mucho

más alta que la última representada: rompe la imagen del diagrama.

Finalmente, aunque lo más frecuente es que el principio de Pareto se cumpla, no siempre es así y podemos encontrarnos con diagramas relativamente planos, en los que todos los criterios tienen niveles parecidos de incumplimientos. En estos casos, la priorización para intervenir, si es que se establece alguna, se basará en otras consideraciones (importancia relativa de cada criterio, facilidad de asumir su mejora, etc.) que no son las puramente cuantitativas.

6. Análisis de los datos de estudios de causas hipotéticas

En los capítulos 5 y 6 vimos dos tipos de estudios dentro de los ciclos de mejora: estudios del nivel de calidad (capítulo 5), en los cuales medimos el nivel de cumplimiento de los requisitos o criterios que de antemano sabemos se relacionan con la calidad (en realidad son la definición operativa de calidad del servicio que se evalúa); y estudios de causas hipotéticas (capítulo 6), cuando lo que medimos es, normalmente, sólo un criterio de calidad y buscamos qué causas (frecuentemente de tipo organizacional) pueden hacer que esté a nive-

les no deseables. Por ejemplo: ¿por qué el tiempo de espera es más largo de lo deseable?, ¿qué factores estructurales y organizacionales influyen en que el diagnóstico de la hipertensión no se haga correctamente?, etc. El segundo tipo de estudios precisa de la medición simultánea del nivel de calidad (reducido normalmente a un solo criterio) y las causas hipotéticas que influyen en la mala calidad.

Como vimos en el capítulo 6, el análisis de los datos de este tipo de estudios consiste esencialmente en ver el grado de asociación entre las diversas causas hipotéticas y el nivel de calidad. Si las causas han sido medidas subjetivamente (por ejemplo mediante encuesta de causas) su análisis puede hacerse mediante un diagrama de Pareto (ver un ejemplo en las figuras 7.1 y 7.2 del capítulo 6); si por el contrario las causas hipotéticas se han medido empíricamente y tenemos datos reales, podemos realizar análisis estadístico y gráfico para evidenciar su asociación con el criterio o indicador de calidad que representa el problema a solucionar y siguiendo la metodología que sea más adecuada según el tipo de variables que hayamos medido. Todo ello fue revisado en el capítulo 6, que debe servir de guía para el análisis y presentación de los datos de este tipo de estudios.

El diagrama de Pareto más completo tiene dos escalas y reúne dos gráficos: la escala y eje de la izquierda. Este último va de 0 hasta el total de incumplimientos de la muestra y sirve de referencia para el gráfico de barras de la frecuencia absoluta (número de incumplimientos) de cada criterio; la escala de 0 a 100 en el eje de la derecha sirve de referencia para representar la curva de frecuencia relativa acumulada. Ambas escalas deben tener altura idéntica.

El diagrama de Pareto asume que todos los tipos de defectos son igualmente importantes.

El principio de Pareto no siempre se cumple. En estos casos, el diagrama no resulta útil para priorizar intervenciones para mejorar.

En los estudios de causas hipotéticas se analiza la relación entre estas causas y el criterio o indicador de calidad que representa el problema analizado.

La cuantificación subjetiva de las causas mediante encuesta a los implicados puede analizarse con diagrama de Pareto.

BIBLIOGRAFÍA

- Briones G. Métodos y técnicas de investigación para las ciencias sociales. Distrito Federal: Trillas, 1990. Cap. 8 y 9.
- Farnum NR. Modern Statistical quality control and improvement. Belmont: Duxbury Press, 1994. Cap. 4.
- Juran JM. Conceptos básicos. En: Juran JM, Gryna FN, Bingham RS. Manual de control de la calidad. 2ª ed. Barcelona: Reverté, 1990. Cap. 2.
- Richards LE, LaCava JJ. Business Statistics. New York: McGraw Hill, 1983. Cap. 7.
- Wadsworth HM, Stephens KS, Godfrey AB. Modern methods for quality control and improvement. New York: John Wiley and sons, 1986. Cap. 9.

Anexo 7.1

Intervalo (i) para un grado de confianza del 95%, según estimación puntual y tamaño muestral

Estimación puntual (%)	Tamaño muestral (n)					Tamaño muestral (n)					
	30	50	60	100	200	30	50	60	100	200	
1	85.0	2.8	2.5	2.0	1.4	51	17.9	13.9	12.6	9.8	6.9
2	5.0	3.9	3.5	2.7	1.9	52	17.9	13.8	12.6	9.8	6.9
3	6.1	4.7	4.3	3.3	2.4	53	17.9	13.8	12.6	9.8	6.9
4	7.0	5.4	5.0	3.8	2.7	54	17.8	13.8	12.6	9.8	6.9
5	7.8	6.0	5.5	4.3	3.0	55	17.8	13.8	12.6	9.8	6.9
6	8.5	6.6	6.0	4.7	3.3	56	17.8	13.8	12.6	9.7	6.9
7	9.1	7.1	6.5	5.0	3.5	57	17.7	13.7	12.5	9.7	6.9
8	9.7	7.5	6.9	5.3	3.8	58	17.7	13.7	12.5	9.7	6.8
9	10.2	7.9	7.2	5.6	4.0	59	17.6	13.6	12.4	9.6	6.8
10	10.7	8.3	7.6	5.9	4.2	60	17.5	13.6	12.4	9.6	6.8
11	11.2	8.7	7.9	6.1	4.3	61	17.5	13.5	12.3	9.6	6.8
12	11.6	9.0	8.2	6.4	4.5	62	17.4	13.5	12.3	9.5	6.7
13	12.0	9.3	8.5	6.6	4.7	63	17.3	13.4	12.2	9.5	6.7
14	12.4	9.6	8.8	6.8	4.8	64	17.2	13.3	12.1	9.4	6.7
15	12.8	9.9	9.0	7.0	4.9	65	17.1	13.2	12.1	9.3	6.6
16	13.1	10.2	9.3	7.2	5.1	66	17.0	13.1	12.0	9.3	6.6
17	13.4	10.4	9.5	7.4	5.2	67	16.8	13.0	11.9	9.2	6.5
18	13.7	10.6	9.7	7.5	5.3	68	16.7	12.9	11.8	9.1	6.5
19	14.0	10.9	9.9	7.7	5.4	69	16.6	12.8	11.2	9.1	6.4
20	14.3	11.1	10.1	7.8	5.5	70	16.4	12.7	11.6	9.0	6.4
21	14.6	11.3	10.3	8.0	5.6	71	16.2	12.6	11.5	8.9	6.3
22	14.8	11.5	10.5	8.1	5.7	72	16.1	12.4	11.4	8.8	6.2
23	15.1	11.7	10.6	8.2	5.8	73	15.9	12.3	11.2	8.7	6.2
24	15.3	11.8	10.8	8.4	5.9	74	15.7	12.2	11.1	8.6	6.1
25	15.5	12.0	11.0	8.5	6.0	75	15.5	12.0	11.0	8.5	6.0
26	15.7	12.2	11.1	8.6	6.1	76	15.3	11.8	10.8	8.4	5.9
27	15.9	12.3	11.2	8.7	6.2	77	15.1	11.7	10.6	8.2	5.8
28	16.1	12.4	11.4	8.8	6.2	78	14.8	11.5	10.5	8.1	5.7
29	16.2	12.6	11.5	8.9	6.3	79	14.6	11.3	10.3	8.0	5.6
30	16.4	12.7	11.6	9.0	6.4	80	14.3	11.1	10.1	7.8	5.5
31	16.6	12.8	11.7	9.1	6.4	81	14.0	10.9	9.9	7.7	5.4
32	16.7	12.9	11.8	9.1	6.5	82	13.7	10.6	9.7	7.5	5.3
33	16.8	13.0	11.9	9.2	6.5	83	13.4	10.4	9.5	7.4	5.2
34	17.0	13.1	12.0	9.3	6.6	84	13.1	10.2	9.3	7.2	5.1
35	17.1	13.2	12.1	9.3	6.6	85	12.8	9.9	9.0	7.0	4.9
36	17.2	13.3	12.1	9.4	6.7	86	12.4	9.6	8.8	6.8	4.8
37	17.3	13.4	12.2	9.5	6.7	87	12.0	9.3	8.5	6.6	4.7
38	17.4	13.5	12.3	9.5	6.7	88	11.6	9.0	8.2	6.4	4.5
39	17.5	13.5	12.3	9.6	6.8	89	11.2	8.7	7.9	6.1	4.3
40	17.5	13.6	12.4	9.6	6.8	90	10.7	8.3	7.6	5.9	4.2
41	17.6	13.6	12.4	9.6	6.8	91	10.2	7.9	7.2	5.6	4.0
42	17.7	13.7	12.5	9.7	6.8	92	9.7	7.5	6.9	5.3	3.8
43	17.7	13.7	12.5	9.7	6.9	93	9.1	7.1	6.5	5.0	3.5
44	17.8	13.8	12.6	9.7	6.9	94	8.5	6.6	6.0	4.7	3.3
45	17.8	13.8	12.6	9.8	6.9	95	7.8	6.0	5.5	4.3	3.0
46	17.8	13.8	12.6	9.8	6.9	96	7.0	5.4	5.0	3.8	2.7
47	17.9	13.8	12.6	9.8	6.9	97	6.1	4.7	4.3	3.3	2.4
48	17.9	13.8	12.6	9.8	6.9	98	5.0	3.9	3.5	2.7	1.9
49	17.9	13.9	12.6	9.8	6.9	99	3.6	2.8	2.5	2.0	1.4

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

p: estimación puntual
n: tamaño

Si $\frac{n}{N} < 0,1$

$$i' = i \cdot \sqrt{\frac{N-n}{N-1}}$$

Anexo 7.2**Indicaciones para la aplicación práctica de los contenidos de este capítulo**

Las indicaciones y formatos que siguen están orientados a presentar y analizar los resultados de la evaluación realizada según los diseños de estudio propuestos en los capítulos 5 y 6. Se subraya que las estimaciones de cumplimiento de los criterios deben darse con sus correspondientes intervalos de confianza y que en el análisis de los defectos de calidad o incumplimientos de los criterios evaluados, o en su caso, de las causas hipotéticas, es útil emplear un diagrama de Pareto completo (dos ejes, diagrama de barras, curva acumulada y punto de corte de los pocos vitales).

Estimación del grado de cumplimiento de los criterios evaluados

Si se han evaluado todos los casos (censo o universo de casos) en vez de una muestra, el grado de cumplimiento real del criterio (C) se desprende de la siguiente fórmula:

$$C = \frac{N_c}{N}$$

N_c : número de cumplimientos

N : tamaño del marco muestral

Para expresarlo como "porcentaje" se multiplica por 100

En el caso, más frecuente, de que la evaluación se haya realizado en una parte representativa del conjunto de los casos (una muestra), el cumplimiento real del criterio (C) no se puede conocer con exactitud debido a que el propio proceso de muestreo ha introducido un error en la medición.

Sin embargo, sí se puede establecer, con una determinada probabilidad (grado de confianza), el intervalo de valores en el que se encuentra comprendido el cumplimiento real del criterio. Sus límites (máximo y mínimo) se pueden determinar fijando previamente la confianza del intervalo y calculando posteriormente: (i) la estimación puntual del nivel de cumplimiento y (ii) el error estándar de la medición. Este procedimiento se desarrolla en las siguientes etapas y es el que se deberá seguir cuando se hayan evaluado muestras.

Paso 1º. Para estimar el nivel de cumplimiento puntual (P) y darlo como porcentaje, la fórmula que se utiliza es análoga a la anterior cuando la muestra se ha obtenido mediante un procedimiento sistemático, aleatorio simple o estratificado proporcional:

$$P = \frac{n_c}{n} \cdot 100$$

n_c : número de cumplimientos en la muestra

n : tamaño de la muestra

Pero si se ha efectuado un muestreo estratificado no proporcional, se debe realizar un cálculo ponderado en base al peso relativo de cada estrato, aplicando la siguiente fórmula:

$$P = \frac{1}{N} \sum N_h \cdot P_h$$

N : Tamaño del universo de estudio

N_h : Tamaño de cada estrato.

P_h : Proporción de cumplimiento en la muestra de cada estrato

Como siempre, para expresarlo como porcentaje lo multiplicaremos por 100

Continúa.../

/... continuación

Paso 2º. Una vez que se ha realizado la estimación puntual debe tomarse una decisión en relación al grado de confianza; con él se va a determinar el intervalo. Habitualmente se consideran como valores aceptables los que son iguales o superiores al 95%. A esta confianza le corresponde un valor z (variable tipificada de la curva normal) de 1.96. Valores correspondientes a otros valores de confianza utilizados habitualmente se encuentran también en el cuadro 7.1 de este capítulo.

Paso 3º. Calcular el error estándar o desviación estándar de la estimación puntual y el correspondiente intervalo para la confianza elegida (i). Aunque las operaciones de cálculo son bastante sencillas, se pueden simplificar utilizando una hoja de cálculo o una calculadora programable, especialmente cuando se han evaluado varios criterios. Las fórmulas, dependiendo del muestreo empleado, son:

Muestreo aleatorio simple o sistemático:

$$i = z \cdot \sqrt{\frac{p(1-p)}{n}}$$

Muestreo aleatorio estratificado proporcional:

$$i = z \cdot \sqrt{\frac{1-f}{n} \sum W_h \frac{n_h}{n_h-1} P_h (1-P_h)}$$

Muestreo aleatorio estratificado no proporcional:

$$i = z \cdot \sqrt{\sum W_h^2 (1-f_h) \frac{P_h (1-P_h)}{n_h-1}}$$

N : tamaño de la muestra

N : tamaño del universo o marco muestral.

h : número de estratos

n_h : tamaño de la muestra en cada estrato

N_h : tamaño de cada estrato

Σ : sumatorio de todos los estratos.

P_h : proporción de cumplimientos de cada estrato

f_h : fracción de muestreo en cada estrato.

f : fracción de muestreo $f = \frac{n}{N}$

W_h : proporción de casos de cada estrato (N_h) en el total del universo muestral:

$$W_h = \frac{N_h}{N}$$

Si hemos decidido una confianza del 95%, el valor de z será 1.96 en todas las fórmulas.

Paso 4º. Reajustar el intervalo en función del tamaño del universo si $\frac{n}{N} > 0.1$, y el muestreo empleado ha sido aleatorio simple o sistemático mediante la fórmula siguiente:

$$i \text{ (ajustada)} = i \cdot \sqrt{\frac{N-n}{N-1}}$$

Paso 5º. Establecer el valor máximo y el valor mínimo (intervalo de confianza [IC]) dentro del cual se encuentra el grado de cumplimiento real del criterio con la probabilidad (confianza) fijada en el paso 2º (habitualmente del 95%).

$$IC_{(95\%)} = P \pm i$$

Análisis de los defectos de la calidad y priorización de la intervención. Diagrama de Pareto

Elaboración de la tabla de frecuencias absolutas y relativas

- Paso 1º.** Diseñar una tabla con el listado de los criterios y, para cada uno de ellos, su frecuencia absoluta y relativa de incumplimientos. Esta última se calcula dividiendo la frecuencia absoluta de incumplimientos del criterio en cuestión entre el total de incumplimientos hallados en la muestra (suma de incumplimientos de todos los criterios). En caso de haber medido causas hipotéticas, se trata cada una de ellas como un criterio. Si se ha hecho una encuesta de causas, se considera el número de casos que cada causa se valora como muy importante.
- Paso 2º.** Clasificar los criterios de la tabla en orden decreciente al número de incumplimientos.
- Paso 3º.** Calcular la frecuencia relativa acumulada sumando al porcentaje relativo de incumplimientos de cada criterio los incumplimientos que suponen todos los anteriores a él en la tabla.

Elaboración del gráfico

- Paso 4º.** Representar mediante barras en el eje de abscisas (eje x) de un plano cartesiano los criterios empleados en la evaluación. La altura de cada barra corresponde a las frecuencias absoluta y relativa de incumplimientos, tomando como referencia los ejes verticales (y) izquierdo y derecho respectivamente. Ambos ejes alcanzan la misma altura correspondiendo el izquierdo al total de incumplimientos y el derecho al porcentaje relativo (0-100%).
- Paso 5º.** Dibujar la curva de frecuencia relativa acumulada.
- Paso 6º.** Señalar, en la curva anterior, el punto que corresponde al 60-80% de los incumplimientos encontrados y comprobar si se cumple la ley de los pocos vitales y muchos triviales, en cuyo caso, la priorización de la intervención se dirigirá a los criterios comprendidos dentro de los pocos vitales.

Nota: si se han estudiado causas hipotéticas, cada una de ellas es como si fuera un criterio. Si se ha estudiado una encuesta de causas, se grafica la frecuencia con que se da a cada una la valoración de "muy importante".

Capítulo 8

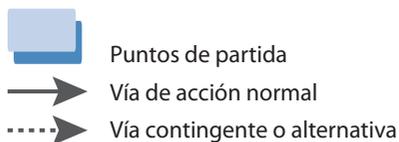
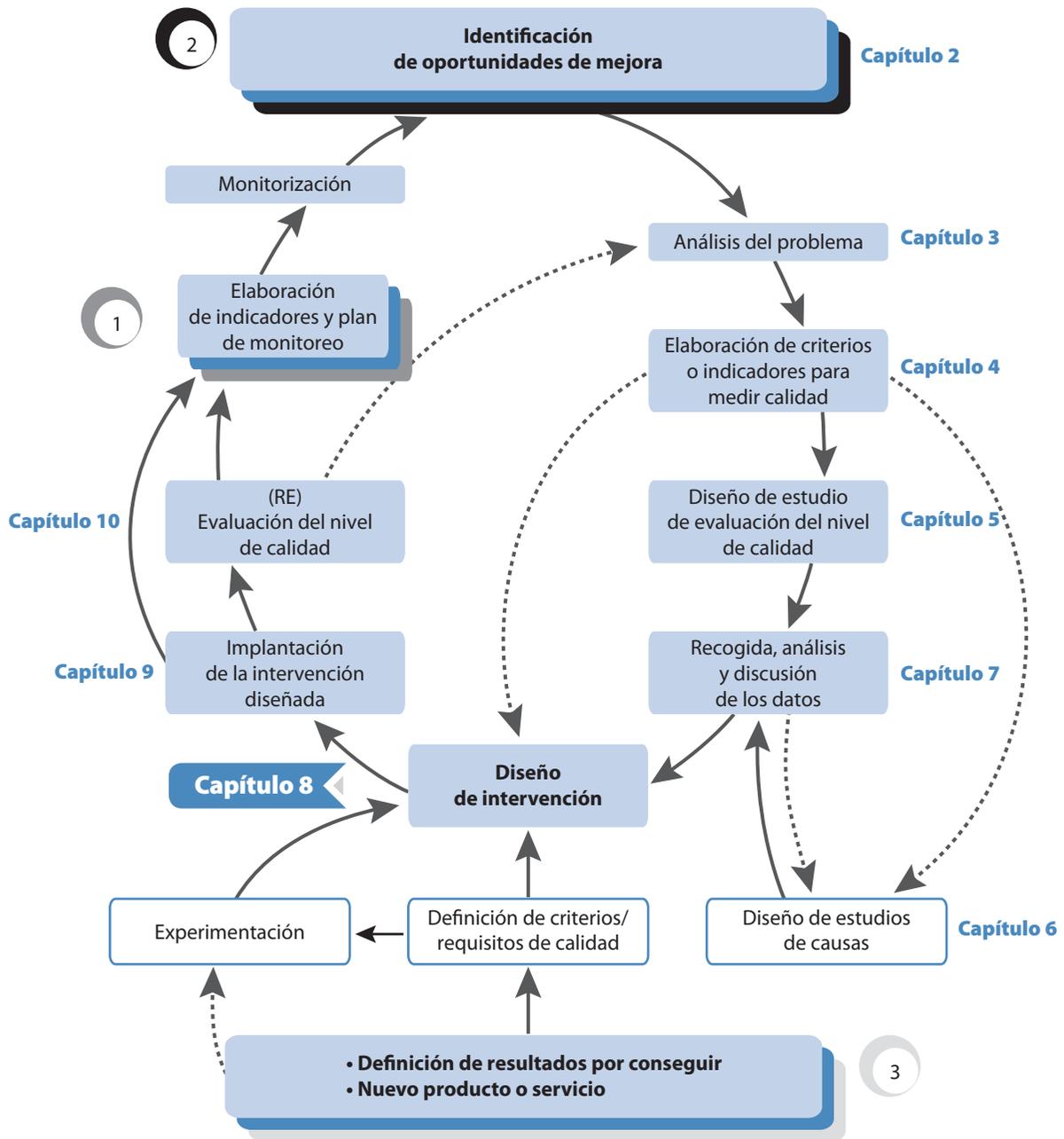
Diseño de intervenciones para mejorar

Contenido general y objetivos específicos

1. Contextualización
2. Diseño de intervenciones para mejorar. Directrices generales
3. Métodos para un diseño participativo

Anexo 8.1. Indicaciones para la aplicación práctica de los contenidos de este capítulo

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Contenido general

Saber qué hacer para mejorar los aspectos evaluados y diseñar su puesta en práctica es el objetivo central de los ciclos de mejora. Este capítulo desarrolla este paso esencial del ciclo, haciendo énfasis en la importancia de basar las intervenciones en los datos obtenidos y en la necesidad de hacer partícipe en su diseño a todos los que vayan a estar implicados en su puesta en práctica. Junto a las directrices generales sobre estos temas, se describen metodologías (diagrama de afinidades, encuesta de intervenciones, votación múltiple) para realizar diseños participativos de una forma estructurada.

Objetivos específicos

En este capítulo se describen los métodos para:

1. Razonar las intervenciones para mejorar, sobre la base de la información obtenida en la evaluación.
2. Compartir la identificación de soluciones con los compañeros implicados.
3. Conducir la construcción de un diagrama de afinidades.
4. Construir un diagrama de afinidades.
5. Conducir un grupo de votación múltiple.
6. Resumir los resultados de un grupo de votación múltiple.
7. Describir en términos operativos la intervención propuesta para mejorar el problema evaluado.

*¡Datos, datos, datos!- gritó impacientemente-. ¡No puedo hacer ladrillos sin arcilla!”
Sherlock Holmes*

1. Contextualización

Después de analizar el problema de calidad (capítulo 3), decidir con qué criterios vamos a medir el nivel a que nos encontramos (capítulo 4) y, sobre todo, tener datos que analizar y discutir (capítulo 7), debemos decidir qué hacer para mejorar y poner en práctica las acciones de mejora que se hayan acordado.

Para estos últimos objetivos, diseñar y poner en práctica la intervención para mejorar, basta en ocasiones con discutir qué es lo que hay que mejorar con los implicados en el proceso que se ha revelado defectuoso o con nivel de calidad mejorable. Sin embargo, al diseñar la intervención, también podemos utilizar métodos estructurados que aseguren la participación de los implicados en implementarla, así como formas de realizar un seguimiento explícito de su puesta en marcha. En este capítulo se explica una selección de estos métodos, eficaces pero de aplicación sencilla, para realizar un diseño participativo de la intervención, y en el siguiente capítulo las relativas al seguimiento de la implementación. De entre los métodos útiles para un diseño participativo vamos a destacar la encuesta de intervenciones, la votación múltiple y, sobre todo, el diagrama de afinidades por su sencillez y versatilidad de aplicación en éste, pero también en otros pasos del ciclo de mejora.

2. Diseño de intervenciones para mejorar. Directrices generales

La principal virtud que pueden tener las acciones que tomemos para mejorar es precisamen-

te eso: que sean efectivas para elevar el nivel de calidad.

Podríamos enumerar un listado más o menos extenso de características de la intervención para mejorar, de forma que nos aseguremos que la efectividad sea más probable; características como que sea realista, factible, aceptada y consensuada, válida, etc. Un listado que podría ser semejante al que vimos para definir los criterios de calidad en el capítulo 4 (de hecho, los criterios de calidad también los definimos como objetivos de calidad que queríamos cumplir o tener). Sin embargo, vamos a seleccionar dos requisitos que son de especial relevancia en el diseño de intervenciones para mejorar: (i) que se base siempre que sea posible en datos y su análisis; y (ii) que se diseñen y acuerden de forma participativa, con la implicación de quienes han de ponerlas en práctica.

2.1. Diseño de intervención basada en datos

Hemos visto en los capítulos anteriores que tanto la identificación de oportunidades de mejora como el análisis de sus causas pueden realizarse sin disponer de datos, utilizando exclusivamente el conocimiento directo que se pueda tener de los servicios que se consideran. De igual manera, es posible decidir, sin realizar medición alguna, cuáles son las causas principales de la calidad deficiente (sobre todo cuando hablamos de causas hipotéticas, no de criterios o requisitos de calidad) y las intervenciones que debemos poner en marcha para eliminarlas. Aunque siempre va a ser necesario medir al menos un criterio o indicador para saber el punto de partida y si hemos logrado

El diseño de la intervención para mejorar se basa principalmente en el análisis y discusión de los resultados de la evaluación, pero hay métodos estructurados para asegurarse que sea participativo y hacer seguimiento de su implementación. En este y en el siguiente capítulo se revisa una selección de estos métodos.

Los estudios de causas hipotéticas se realizan con cierta frecuencia sin medición directa y la intervención se basa en el conocimiento que se tiene sobre el proceso analizado. No así los estudios de nivel de calidad que habitualmente cuantifican el nivel de cumplimiento de los criterios diseñados.

Las intervenciones son más efectivas si se basan en datos y se diseñan con la participación de los implicados en ponerlos en práctica.

o no mejorar tras la intervención, se utilizan métodos estructurados para el consenso sobre causas e intervenciones como el *diagrama de causa y efecto* y el *análisis de campos de fuerza* (ya vistos en capítulos anteriores) y el *diagrama de afinidades* (que veremos más adelante). Dichos métodos en ocasiones se utilizan sin que vengán precedidos o acompañados de medición con datos reales de las suposiciones que los sustentan. Esto es más frecuente a la hora de decidir cuáles son las causas hipotéticas (normalmente estructurales, organizativas, propias de un contexto concreto) de que los criterios o requisitos de calidad no se cumplan.

Si queremos ser estrictos y rigurosos, antes de intervenir deberíamos realizar una medición de estas causas hipotéticas y un estudio estadístico de su asociación o no al problema de calidad considerado, tal y como se explica en el esquema del capítulo 6.

Sin embargo, si lo que hemos realizado es un estudio del nivel de calidad, es decir, del nivel de cumplimiento de una serie de criterios o requisitos de calidad, hacemos equivalente criterio incumplido a causa de calidad deficiente, y lo que hemos de hacer es ver qué criterios y en qué medida se incumplen, tal como hemos visto en el capítulo 7; reflexionaremos sobre la forma en que podamos hacer que se cumplan. En ocasiones, no habitualmente, puede ser necesario realizar un estudio de causas hipotéticas del incumplimiento de un determinado criterio para encontrar la mejor forma de intervenir sobre él. Esta opción está representada en el esquema gráfico introducido que aparece al principio de cada capítulo, pero lo más importante es la reflexión cualitativa estructurada en torno a cómo mejorar los criterios que no se cumplen.

2.2. Medición de las causas y calidad de los registros

Otra cuestión diferente y que aparece con frecuencia en la discusión sobre los niveles de cumplimiento de los criterios de calidad, si éstos han sido medidos en historias clínicas u otros documentos de registro de actividades realizadas, es el tema de si los datos que tenemos miden la calidad de la atención o la calidad de los registros o historias clínicas: el argumento más

frecuente es “yo lo hago correctamente pero no tengo tiempo de registrarlo”. ¿Cómo reaccionar ante ello?, ¿puede ser cierto que no sea la calidad de la atención lo que hayamos evaluado? Al realizar cualquier medición estamos expuestos a una serie de errores dependientes de la fuente de datos utilizada, del instrumento de medida y del propio método de medición; sin embargo, todo ello debe estar suficientemente controlado y no justifica una descalificación total del tipo de la que se suele argumentar. Este argumento se contrarresta por sí solo si los criterios que hemos medido son suficientemente relevantes, de forma que la historia clínica (y lo que en ella se anote) más que una fuente de datos es, para esos criterios, un instrumento necesario para la correcta atención al problema de salud que se evalúa. Esta característica se da claramente cuando ocurre uno de los casos siguientes (resumidos en el cuadro 8.1):

(i) El dato que buscamos en la historia forma parte indiscutible del proceso asistencial, de modo que su ausencia condiciona actuaciones posteriores y/o la interpretación de otros datos del proceso. Es el caso, por ejemplo, de los contenidos de las visitas de control, los resultados de exploraciones realizadas o los tratamientos prescritos. ¿Para qué se hacen los controles sino sirven para ver una evolución?, ¿cómo vamos a saber qué se ha prescrito con anterioridad a la hora de analizar otras alternativas, si fiamos este dato exclusivamente a nuestra memoria y la del paciente? Esto es atención defectuosa.

(ii) El dato que buscamos en la historia lo sabemos necesario para la valoración del paciente y antes o después lo vamos a indagar, de forma que si lo indagamos y no lo anotamos, corremos el riesgo de repetir (nosotros mismos u

A veces se desconfía de los datos que se han extraído de historias clínicas u otros sistemas de registro. La mejor manera de prevenir este problema es una selección adecuada de los criterios, de forma que su anotación forme parte claramente del propio proceso asistencial evaluado.

Cuadro 8.1

Evaluación con base en datos registrados

Más válida si:

- El registro del dato es importante para el proceso que se evalúa.
- El registro del dato evita repeticiones.

otros profesionales) preguntas y exploraciones que sólo hace falta hacer una vez. Este sería el caso de los antecedentes familiares, personales y toda la serie de exploraciones básicas que hay que realizar en caso de patologías crónicas.

En todos estos casos no vale el argumento “yo lo hago pero no lo anoto” como forma de justificar una buena actuación profesional: la propia no anotación la convierte en actuación defectuosa por su probable repercusión en valoraciones posteriores del mismo paciente, o en la repetición innecesaria de preguntas y exploraciones.

Conviene insistir, sin embargo, en la necesidad de que los datos que se vayan a buscar en la historia clínica u otros registros, estén *perfectamente justificados en función de su interés para el proceso asistencial a valorar*. La historia clínica y documentos semejantes son un medio, no un fin en sí mismos. Puede ser un inconveniente para la credibilidad y justificación de los ciclos de mejora basarlos en lo que se registra o no, sin asegurarse previamente la importancia, para un determinado proceso asistencial, de lo que se va a medir.

En realidad, más que el probable (e injustificable, si los criterios son adecuados) infraregistro, lo que puede resultar problemático para valorar la calidad en base a datos registrados es lo contrario: que se registre como hecho algo que en realidad no se haya hecho. Afortunadamente no parece ser lo más frecuente, pero en algunos estudios observacionales (comparando lo que se observa que se hace con lo que queda registrado) se ha visto que este problema puede existir, sobre todo para aparentar que se cumple con requisitos y protocolos obligatorios. Un caso de este tipo lo hemos visto recientemente en nuestro medio hospitalario en relación a la cumplimentación de la lista de comprobación (*checklist*) quirúrgica. La conclusión sería que ante una evaluación con base en datos registrados, más que cuestionar los cumplimientos bajos (que siempre van a ser problemas de calidad si los criterios están bien elegidos y diseñados), lo único que podría ser cuestionado es encontrar cumplimientos sospechosamente altos.

En cualquier caso, una vez ante los datos analizados y discutidos, debemos proceder a

diseñar cómo actuar para mejorar. Ya hemos apuntado la necesidad de que este diseño sea participativo, incluyendo e implicando a los que están de todas formas implicados en el proceso que vamos a mejorar.

3. Métodos para un diseño participativo

La mera discusión de los datos y de las sugerencias del equipo implicado en relación a las actuaciones para mejorar puede ser suficiente para decidir qué hacer. No obstante, muchas veces es conveniente realizar la discusión de forma estructurada y con un cierto método que garantice un uso eficiente de las aportaciones de todos, a la vez que nos asegura lo más posible su implicación y la sensación de “pertenencia” sobre las actuaciones acordadas.

El cuadro 8.2 contiene una relación de algunos de los métodos que pueden ayudar a realizar este diseño participativo de forma estructurada. Dos de ellos ya han sido vistos con otros fines en otros capítulos (por ejemplo la técnica de grupo nominal en el capítulo 2 como método grupal cualitativo para priorizar problemas de calidad). Si se revisan sus objetivos y funcionamiento, se hace fácilmente evidente su utilidad también para diseñar intervenciones concretas en los ciclos de mejora; sólo hay que formular el objetivo del análisis, haciendo de la propuesta de acciones concretas el objetivo de reflexión para el grupo nominal. Los otros métodos mencionados en el cuadro 8.2, también pueden utilizarse con objetivos diferentes a la búsqueda de intervención para mejorar; por ejemplo la encuesta de causas/intervenciones y la votación múltiple son de utilidad en el análisis de problemas y cuantificación de causas, si no queremos realizar una

Existen varios métodos grupales para realizar un diseño de intervención participativo; aunque no siempre es necesario ni obligatorio utilizar alguno de ellos, vamos a destacar la utilidad para este fin del Diagrama de Afinidades.

Cuadro 8.2

Métodos para un diseño de intervención participativo

- Análisis de campos de fuerza
- Técnica de grupo nominal
- Encuesta de causas y/o intervenciones
- Votación múltiple
- Diagrama de afinidades

El diagrama de afinidades es una técnica grupal que consiste en ordenar las intervenciones y convertirlas en líneas de actuación y es útil para recoger todas las ideas del grupo para mejorar el problema evaluado en un plan de acción. Para ello hay que seguir unos pasos perfectamente estructurados

medición empírica, tal como vimos en el capítulo 6. Probablemente el diagrama de afinidades es, junto a la encuesta de causas/intervenciones, uno de los métodos más genuinamente útiles para diseñar estrategias e intervenciones de forma participativa, y es conveniente destacarlos sobre los demás. Vamos a recordar brevemente en qué consiste la encuesta sobre causas y la votación múltiple, y revisar más en detalle el diagrama de afinidades.

3.1. Encuesta sobre intervenciones para mejorar

Tanto el formato de las preguntas como la escala de valoración pueden ser de varios tipos, pero finalmente lo que se persigue es tener una cuantificación/priorización de las diversas acciones en relación a su importancia para resolver el problema. El resultado de la encuesta puede presentarse y analizarse con un diagrama de Pareto, en donde se subrayen y prioricen las actividades que han de componer la intervención en función de la valoración recibida.

En el capítulo 6 se explica con un ejemplo, relacionado con la investigación de causas que conviene revisar, puesto que la utilidad de este método es similar para priorizar intervenciones. El procedimiento es idéntico cuando lo que queremos es priorizar intervenciones. La única diferencia es que en vez de preguntar sobre las posibles causas, se pregunta sobre las posibles intervenciones.

3.2. Votación múltiple

La votación múltiple es también un método para seleccionar las intervenciones más importantes de la lista que se haya elaborado. Consiste en una serie de votaciones entre los implicados en el proceso a mejorar, cada una de las cuales reduce la lista hasta que quedan seleccionadas aquellas sobre las que basar la estrategia de mejora. Los pasos a seguir son los mismos que cuando se utiliza para priorizar causas, solo que el listado de partida es un listado de acciones vez de causas hipotéticas.

La votación múltiple no deja de ser una técnica de consenso que tiene el riesgo, como todas las técnicas que no manejan datos, de

equivocarse si las percepciones del grupo no son las correctas. La reevaluación de la calidad con base en los criterios o el indicador que representa el problema y su comparación con la situación (medición) de partida va a ser la única forma de saber si el grupo estuvo o no acertado.

3.3. Diagrama de afinidades

La construcción de un diagrama de afinidades es un proceso grupal que comienza con la generación de ideas, opiniones, o de actividades sobre el tema que se propone, y que termina con esas ideas o actividades ordenadas de forma lógica en grupos afines que puedan constituir los elementos de la estrategia a emprender y las actividades a realizar. Es posible combinar el diagrama de afinidades con otros métodos estructurados, como por ejemplo la encuesta de intervenciones, realizando el diagrama como complemento de estos otros métodos, pero también puede ser utilizado independientemente para ordenar y dar forma a las ideas que el grupo tenga sobre las intervenciones a realizar.

La generación de ideas para el diagrama de afinidades es similar a la lluvia o tormenta de ideas (*brainstorming*) en relación a la inexistencia de límites en su número, y se parece a la técnica del grupo nominal en el que cada individuo del grupo debe de escribir las suyas, de forma individual, antes de que se discutan o se comparen con las expresadas por los demás. Los pasos a seguir son los siguientes (cuadro 8.3).

Cuadro 8.3
Diagrama de afinidades

Pasos:

1. Planteamiento del problema.
(Una pregunta: ej.: ¿Qué intervención?)
2. Cada miembro del grupo anota acciones concretas.
(Regla de las tres palabras)
3. Todas las acciones se ponen en el tablón sin orden concreto.
4. Agrupación de las acciones en líneas estratégicas.
5. "Nombrar" los grupos de acciones o líneas estratégicas.
(Unir las acciones parecidas, considerar las acciones "huérfanas")
6. Describir el resultado.

Los métodos de consenso que no manejan datos tienen el riesgo de equivocarse si las percepciones del grupo no son las correctas.

1. El coordinador del grupo que va a realizar la actividad plantea el problema, cuestión o pregunta que se quiere resolver con esta técnica. Por ejemplo, a la vista de los resultados concretos que hemos obtenido en la evaluación del nivel de calidad ¿qué podemos hacer para mejorar?

2. Cada miembro del grupo da sus respuestas a la pregunta propuesta. Estas respuestas no se dan abiertamente sino que se escribe cada una de ella en un papel pequeño (después hay que examinarlas todas a la vez y reagruparlas junto con las de los demás). Para esta actividad hay dos recomendaciones que facilitan los pasos posteriores del diagrama:

(i) Las respuestas que se den no han de ser muy largas sino escuetas; pero tampoco tan escuetas como para que no esté claro qué queremos decir; algunos autores sugieren que se siga la “regla de las tres palabras”, según la cual, cada idea debe exponerse con un mínimo de tres palabras que incluyan preferentemente un sujeto, un verbo y su predicado. Por ejemplo, si pensamos que hay que formarse sobre alguna cuestión concreta no se debe escribir como idea “formación” sino formación sobre qué y a quién. Supongamos, por ejemplo, que el criterio sobre exploración de fondo de ojo en el hipertenso no se cumple porque nunca nadie nos enseñó a hacerlo, y pensamos que debe hacerlo el médico; la actividad o idea que vamos a proponer es “formación sobre exploración de fondo de ojo a los médicos del centro”. Si pensamos que el problema es que el sistema de registro es deficiente y no nos recuerda o invita a realizar la anamnesis y exploración completa que debemos hacer, la idea sería “modificar la estructura del sistema de registro”; o podemos explicitar aún más qué tipo de modificación creemos que es importante, como por ejemplo “añadir casillas con respuesta “sí” y “no” para la anamnesis sobre tabaco y alcohol”, porque podemos pensar que de este modo, al tener forzosamente que señalar que no, desaparece la ambigüedad de pensar que cuando estén en blanco quiere decir que no fuma o no bebe, además de que nos da pie para completar la

información sobre hábitos tóxicos (tabaco y alcohol) si señalamos la casilla del “sí”.

(ii) Utilizar papel autoadhesivo, fácil de pegar y despegar, para realizar posteriormente las agrupaciones y reagrupaciones tanto si se realizan sobre una mesa, como si se realizan sobre un tablón.

En este paso no hay discusión de grupo, cada uno individualmente escribe tantas notas como ideas tiene para proponer.

(iii) Todas las ideas o propuestas de acciones generadas se ponen encima de la mesa o en un tablón sin orden concreto; estando todas a la vista, se puede proceder a su comparación y ordenación.

(iv) Agrupación de las acciones en grupos lógicos, *afines* o relacionados. Si el grupo de personas que está contribuyendo al diagrama es numeroso, se toman turnos de tres o cuatro personas que realizan las agrupaciones de ideas durante un cierto tiempo (por ejemplo, 5 minutos) y luego dan paso a otros miembros del grupo; si el grupo es pequeño (tres a cinco personas) puede realizarse esta actividad uno a uno. Es importante que sea sólo la persona o grupo que esté realizando en cada momento la agrupación de las propuestas las únicas que opinen, discutan o actúen sobre la ordenación, hasta que finalmente queden los grupos de ideas establecidos. Hay una serie de recomendaciones a tener en cuenta para que éste paso, probablemente el más importante de todo el proceso, se realice con la mayor eficiencia:

- Procurar no forzar una idea dentro de un grupo con el que no esté claramente relacionada. Por ejemplo, las actuaciones de formación no tienen nada que ver con las modificaciones de los registros, pero pueden relacionarse con otras acciones que tengan que ver con el personal y las actividades que tengan que realizar.
- Puede haber ideas “huérfanas” no relacionadas con ninguna otra: dejarlas como tales.
- Si se produce un desacuerdo irreconciliable sobre dónde agrupar una determinada acción, puede duplicarse y colocarse en más de un grupo.

El diagrama de afinidades produce un consenso sobre el plan de acción de forma fácil y rápida, y puede completarse con algún instrumento para asegurar y supervisar la implementación del plan de acción.

Esta agrupación de ideas o acciones para mejorar en grupos afines, es lo que da el nombre a esta técnica, cuyo siguiente paso es precisamente explicitar cuáles son los grupos de acciones que hemos identificado.

- (v) Nombrar los grupos de acciones o líneas estratégicas. Una vez que el grupo está de acuerdo en la agrupación y parece que no es posible cambio razonable alguno, se procede a dar nombre a los grupos formados. Pueden resultar, por ejemplo, acciones de formación, reorganización del trabajo, cambios en los sistemas de registro, adquisición de material, etc. Cada una de estas líneas estratégicas contiene actividades concretas que vamos a poner en marcha. Todas ellas configuran la intervención a realizar para mejorar.
- (vi) Describir el resultado de forma clara y ordenada para que todo el grupo y el resto del centro sepan qué es lo que vamos a hacer para mejorar. Esta descripción puede hacerse como diagrama de árbol (cada rama principal es una línea estratégica) o como simple relación de acciones agrupadas. El cuadro 8.4 es un ejemplo de diagrama de afinidades, resultado de una reflexión grupal sobre qué hacer para mejorar la detección, diagnóstico y estudio inicial del hipertenso.

Las ventajas de explicitar las acciones a tomar con un diagrama de afinidades son varias. En primer lugar implican a todo el grupo. En segundo lugar, son normalmente fáciles de hacer y, aunque se propongan muchas ideas, se alcanza el consenso fácil y rápidamente. Adicio-

Cuadro 8.4
Ejemplo de diagrama de afinidades
¿Qué hacer para mejorar la detección, diagnóstico y estudio inicial del hipertenso?

1. **Formación del personal médico**
 - Exploración de fondo de ojo
 - Interpretación e informe de ECG
2. **Cambios en los sistemas de registros**
 - Poner casillas si y no para los hábitos tóxicos
 - Incluir explícitamente todos los ítems que hay que explorar
 - Eliminar lo que no sea necesario
 - Incluir recordatorios al principio indicando que un ítem en blanco no quiere decir que es normal sino que no se ha explorado
3. **Organización del trabajo**
 - Establecer circuitos y responsabilidades de médicos y enfermería para la detección de hipertensos
 - Establecer sistemas específicos de cita y recordatorios para los pacientes en proceso de diagnóstico inicial
 - Establecer algoritmo de procedimientos a realizar para incluir al paciente en programa

nalmente, la participación no verbal que implica, facilita la participación de todo el mundo, incluso de los que habitualmente no lo hacen. Finalmente, queda claro para todo el esbozo de plan de acción que puede completarse asignando responsabilidades concretas para cada acción y utilizando alguno de los métodos para asegurar y supervisar su puesta en marcha, como veremos en el siguiente capítulo.

BIBLIOGRAFÍA

- Goldfield N, Pine M, Pine J. Techniques for analyzing processes and managing team decisions. In: Measuring and managing health care quality. Aspen, 1995. Vol I, Cap. 5.
- Leebow W, Ersoz CJ. Tools for targeting improvements. In: The health care manager's guide to continuous quality improvement. Chicago: AHA, 1991. Cap. 11.

Anexo 8.1**Indicaciones para la aplicación práctica de los contenidos de este capítulo**

Se debe intentar la intervención oportuna para mejorar, a la vista de los resultados de la evaluación del nivel de calidad y/o del estudio de causas hipotéticas. Utilizando los métodos reseñados en este capítulo se pretende que esta intervención se describa de forma estructurada a través de un diagrama de afinidades, a realizar preferentemente en grupo (con los implicados en el problema a resolver) o de forma individual ordenando en grupos afines todas las posibles acciones para mejorar. Para ello:

1. Seguir los pasos para la elaboración de un diagrama de afinidades que se indican en el apartado 3 de este capítulo.
2. Plasmar el resultado en una hoja, con formato semejante al del cuadro 8.4 de este capítulo.

Otra alternativa recomendable es realizar una encuesta de intervenciones y resumir los resultados con un diagrama de Pareto, tal como se muestra en los capítulos 6 y 7.

Capítulo 9

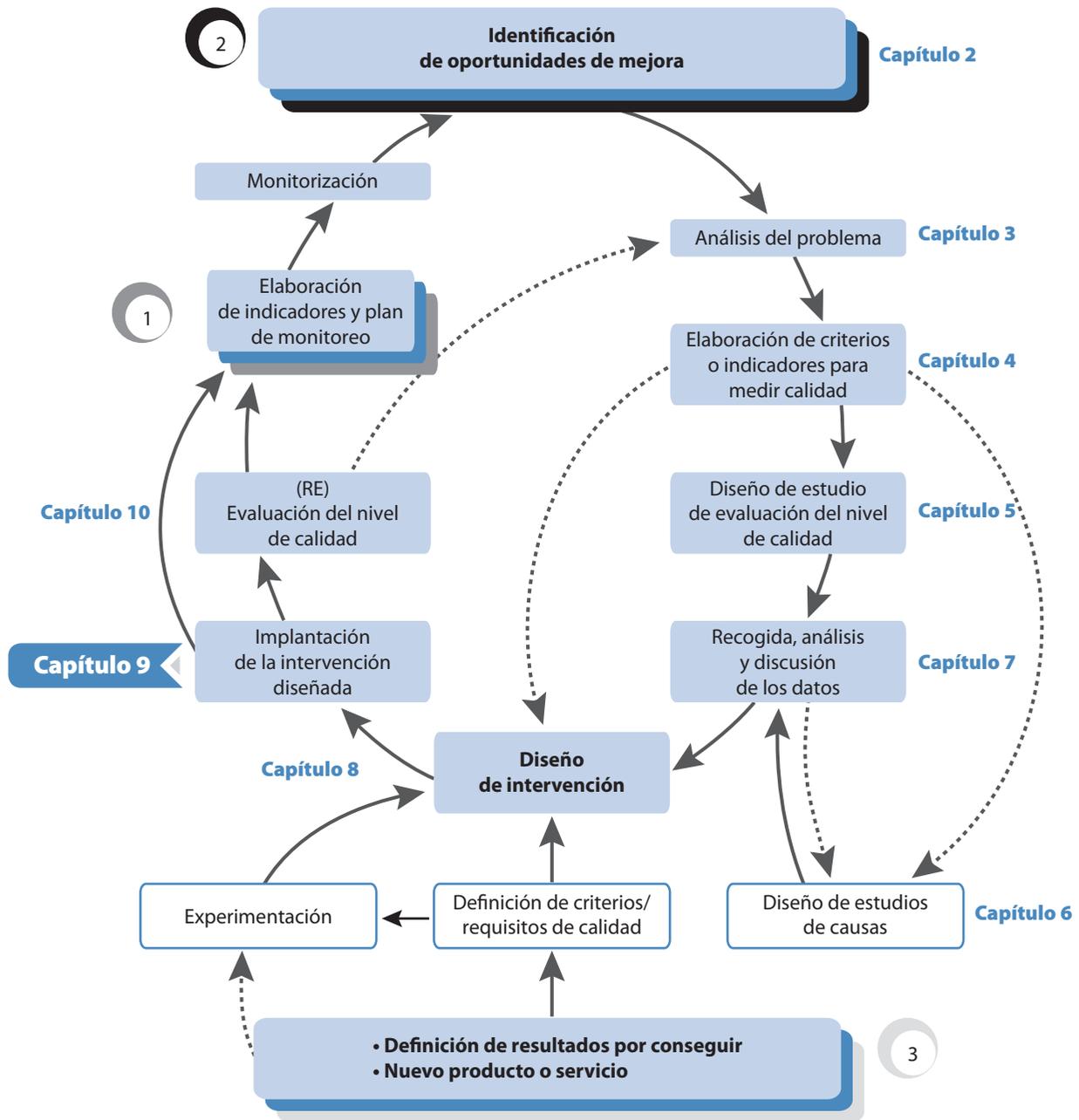
Implementación de intervenciones para la mejora de la calidad. Métodos de seguimiento y control

Contenido general y objetivos específicos

1. Contextualización
2. Métodos de apoyo para la implementación de la intervención para mejorar
3. Seguimiento de la implementación de la intervención para mejorar

Anexo 9.1. Indicaciones para la aplicación práctica de los contenidos de este capítulo

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Puntos de partida

Vía de acción normal

Vía contingente o alternativa

1 Monitorización

2 Ciclo de mejora

3 Diseño

Contenido general

Saber qué hacer para mejorar los aspectos evaluados y diseñar su puesta en práctica es el objetivo central de los ciclos de mejora. Este capítulo complementa al anterior, en el que se muestran metodologías para el diseño participativo de intervenciones para mejorar, revisando métodos y herramientas para planificar y controlar su puesta en práctica (diagrama de Gantt, diagrama de problemas anticipados, carteles narrativos).

Objetivos específicos

En este capítulo se describen los métodos para:

1. Estructurar los componentes y responsabilidades de la intervención propuesta.
2. Construir un Diagrama de Gantt.
3. Desarrollar un Cartel Narrativo del Ciclo de Mejora.
4. Analizar estructurada y explícitamente los problemas que puede tener la puesta en práctica de la intervención diseñada.

Implementación de intervenciones para la mejora de la calidad. Métodos de seguimiento y control

1. Contextualización

Siguiendo los métodos descritos en el capítulo anterior, ha de haber quedado especificada, tanto en términos de líneas estratégicas generales como de acciones específicas a realizar, la intervención a implementar para mejorar el problema evaluado. El siguiente paso es hacer que la intervención diseñada sea una realidad, es decir, implementarla. Asegurarnos la implementación es, en esencia, un ejercicio de planificación operativa en el que tenemos que tener claro cuándo, cómo y quién tiene que hacer cada una de las acciones previstas. En este sentido, cualquiera de las herramientas de utilidad para la planificación operativa nos puede ser de también de utilidad en esta fase del ciclo de mejora. Sin embargo, nos vamos a centrar en las más simples y de uso más común, poniendo énfasis en la conveniencia de utilizar alguna de ellas.

2. Métodos de apoyo para la implementación de la intervención para mejorar

Una vez decidido lo que hay que hacer para mejorar, el plan de acción consecuente puede ser muy simple y no necesitar de herramientas adicionales de planificación operativa. En esos casos, la única recomendación que prevalece de los diversos métodos a utilizar es el uso de un póster o cartel narrativo del proyecto para ir reflejando su progreso y explicitar, a la vista de todos, el compromiso contraído. Para planes de acción complejos, pueden ser de utilidad herramientas de planificación operativa

también complejas como el PERT (Program Evaluation and Review Technique).

*Sin embargo, en la mayoría de los casos, en nuestros ciclos de mejora algunas herramientas más simples como el diagrama de problemas anticipados y el diagrama de Gantt pueden ser suficientes; ambas herramientas se describirán en este texto. En nuestra experiencia, dichas herramientas, aunque recomendables, puede incluso llegar a hacerse necesarias en las propuestas de mejora habituales, sobre todo al principio de los programas o actividades de gestión de la calidad. El cuadro 9.1 contiene la relación de métodos que vamos a revisar.

2.1. Diagrama de Gantt

Los diagramas de Gantt son representaciones de las actividades a realizar en un plan de acción de forma que se visualice su relación temporal desde el principio hasta el fin de la ejecución del plan. El esquema básico es el que

El uso de carteles narrativos y de instrumentos de planificación operativos como el diagrama de Gantt o el de problemas anticipados puede ser de utilidad para el éxito de la intervención para mejorar.

Cuadro 9.1 Métodos comunes para el seguimiento del proceso de mejora	
●	Análisis y vigilancia de problemas anticipados
●	Diagrama de Gantt
●	Carteles narrativos

El diagrama de Gantt representa la secuencia temporal correcta de las actividades del plan de acción. Para un mayor control, se puede dibujar sobre él la marcha real del proyecto.

* Los fundamentos de esta herramienta pueden verse en los siguientes documentos: <http://www.uhu.es/eyda.marin/apuntes/admon/tema9_1.pdf> <http://ocw.uc3m.es/economia-financiera-y-contabilidad/economia-de-la-empresa/material-de-clase-1/PERT.pdf>

se presenta en la figura 9.1. En la parte izquierda del diagrama están las diversas acciones o actividades a realizar ordenadas en secuencia temporal lógica, mientras que en el encabezado (o a veces en la parte inferior) se representa la secuencia de unidades de tiempo que sea adecuada al plan (días, semanas, meses, etc.). El diagrama consiste entonces en señalar con un segmento o una barra horizontal el tiempo de comienzo y final de cada una de las actividades. De esta forma pueden verse las actividades a realizar de forma paralela y las que hay que realizar después de que se completen otras anteriores. También se evidencia el tiempo mínimo de ejecución del plan, la secuencia apropiada de actividades y aquellas que pueden determinar retraso en la aplicación del plan. En este sentido, el diagrama de Gantt es también una forma visualmente más atractiva de presentar un diagrama PERT, otra de las herramientas clásicas de planificación operativa que ya hemos mencionado pero que no vamos a considerar aquí por su mayor complejidad y una menor aplicabilidad en los ciclos de mejora normales.

Esquema para la elaboración de un diagrama de Gantt

En el diagrama de Gantt pueden introducirse algunas modificaciones que lo hacen aún más informativo y útil. Por ejemplo, junto a la columna de tareas o actividades puede añadirse otra en la que se especifican la persona o personas responsables de esa actividad concreta; de igual manera, es más útil como herramienta de seguimiento del proyecto de mejora si se va dibujando la línea de ejecución real de las actividades en paralelo a la establecida en la planificación inicial.

Las intervenciones acordadas pueden plasmarse en uno o más Diagramas de Gantt. A veces conviene tener un diagrama de Gantt para cada una de las líneas estratégicas del diagrama de afinidades. En la figura 9.2 puede verse un ejemplo simple de diagrama de Gantt aplicado a la planificación operativa de los cambios a introducir en los sistemas de registro, una de las líneas estratégicas acordadas en nuestro diagrama de afinidades del cuadro 8.4 del capítulo anterior.

Tareas Actividades	Unidades de Tiempo (días, semanas, meses, etc.)									
	1	2	3	4	5	6	7	8	9	10
1.										
2.										
3.										
4.										
5.										
6.										
7.										
8.										
9.										

Figura 9.1. Esquema para la elaboración de un diagrama de Gantt

Objetivo: Introducción de cambios en los sistemas de registro para el estudio inicial del hipertenso.

Plan de acción (diagrama de Gantt)

Tarea	Responsables	Semana										
		1	2	3	4	5	6	7	8	9	10	
1. Recoger y revisar los documentos existentes	Juan	█										
2. Hacer propuesta de diseño de cambios	Juan y Carmen		█	█	█							
3. Revisar propuesta con todo el equipo	Carmen				█							
4. Pilotar modificaciones acordadas	Rocío y José					█	█	█	█			
5. Discutir resultados	Rocío y José											
6. Aprobar	Juan										█	
7. Imprimir nuevo formato	Miguel											█
8. Sustituir los antiguos impresos por los nuevos	Carmen											
9. Promover y supervisar su utilización	Juan, Rocío y José											█

Figura 9.2. Ejemplo de diagrama de Gantt

2.2. Diagrama de problemas anticipados

Esta herramienta es la plasmación explícita de los probables obstáculos esperables en la puesta en marcha de las acciones acordadas, de forma que se fuerza a pensar también en las posibles soluciones para cuando aparezcan. Es un diagrama que resulta un complemento fácil y lógico para el análisis de campos de fuerza cuando ha sido este el método utilizado para diseñar la estrategia de la intervención para mejorar, dado que los principales problemas esperables se derivan normalmente de las *fuerzas en contra* identificadas en este tipo de análisis. En definitiva

se trata de establecer una planificación *contingente* a la evolución de la puesta en marcha del plan de acción diseñado. El enfoque de base es semejante al de un AMFE (Análisis Modal de Fallos y Efectos) aplicado frecuentemente a la gestión de riesgos en relación a la seguridad del paciente.

La figura 9.3 contiene el esquema básico de este diagrama, cuya práctica es, de todas formas, menos habitual y probablemente también menos efectiva y atractiva que los *carteles narrativos*, la más recomendable y sencilla de las herramientas que sugerimos utilizar, aunque lo más indicado es empezar su construcción a la vez que se inicia el ciclo de mejora.

El diagrama de problemas anticipados sirve como herramienta de planificación contingente a la aparición de obstáculos o problemas presumibles en relación al plan de acción decidido.

Resumen de la estrategia a implantar: (describirla)

¿Qué puede ir mal?	¿Qué probabilidad hay de que ocurra? (a)	¿Qué tan importante es el obstáculo? (b)	Atención prioritaria (a x b)	Acciones preventivas	Acciones para minimizar efectos	Responsables de acción
1.						
2.						
3.						
4.						
5.						
6.						
etc.						

(a) y (b): escala: 1 poco; 2 moderadamente; 3 mucho; 4 muchísimo.

Figura 9.3. Esquema para un diagrama de problemas anticipados

Cualquier actividad de seguimiento formal o informal, es útil y conveniente en la medida que logre que la mejora se consiga.

2.3. Carteles narrativos del proyecto de mejora

El *story board* o *cartel narrativo* del proyecto de mejora es una representación paso a paso del ciclo de mejora. Al utilizarlo estarán los espacios en blanco y se irán ilustrando y llenando a medida que vaya avanzando el proyecto. Para cada paso del ciclo de mejora se va informando de las actividades clave y las herramientas utilizadas. El que esté todo ello expuesto en un lugar visible fomenta la responsabilidad y compromiso del grupo de mejora encargado del proyecto, a la vez que hace posible y fácil recibir comentarios y sugerencias de otros profesionales del centro. La figura 9.4, que ya mostramos en el capítulo 2, es un ejemplo de la estructura de un cartel narrativo. En la fase del ciclo de mejora en que nos encontramos en este capítulo, deberían estar cumplimentados los apartados de *causas principales* y *solución*, además de todos los anteriores del cartel. Si nos parece oportuno, puede añadirse también a continuación de estos dos apartados el Diagrama de Gantt que explicita las acciones a

implementar en la intervención para mejorar, su secuencia temporal y los responsables de las mismas.

3. Seguimiento de la implementación de la intervención para mejorar

Los métodos y herramientas que acabamos de ver tienen como uno de sus principales objetivos el facilitar la continuación y el seguimiento del proyecto de mejora, seguimiento que resulta más fácil si utilizamos alguna de ellas. Sin embargo, también hemos apuntado que muchas veces la intervención para mejorar se decide tras un simple análisis y discusión de los datos de la evaluación, sin que se proceda a ningún tipo de planificación operativa explícita, con actividades, tareas, tiempos y responsabilidades. Aún en ese caso, sería conveniente establecer algún mecanismo de seguimiento o comprobación de que efectivamente, se están haciendo las cosas acordadas. Este seguimiento puede ser algo tan simple como recordatorios o discu-

Narración de EMC

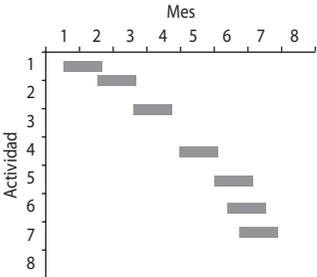
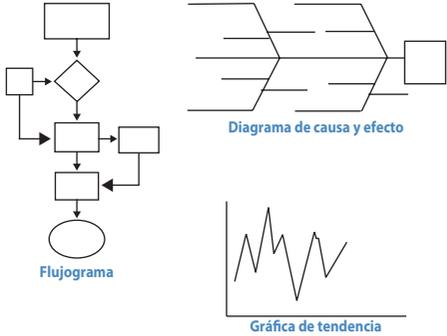
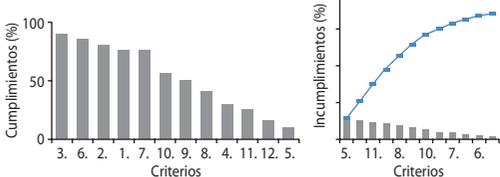
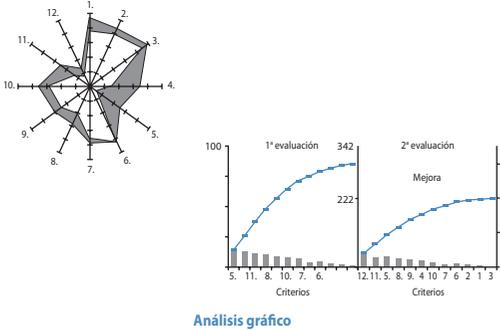
<p>Definición del problema</p>	<p>Equipo</p> <p>Nombres:</p> <p>Roles:</p>	<p>Foto</p> 	<p>Cronograma</p> 																																																																				
<p>Análisis del problema</p> 		<p>Lista de criterios</p> <ol style="list-style-type: none"> 1. Diagnóstico correcto 2. Antecedentes familiares 3. Antecedentes personales 4. Consumo de tabaco 5. Consumo de alcohol 6. Peso y talla y/o IMC 7. Auscultación cardiaca 8. Exploración abdominal 9. Pulsos periféricos 10. Soplos carotídeos 11. Edemas 12. Fondo de ojo <p>Instrumentos</p> <p>Questionario</p> <p>Listado de chequeo</p> <table border="1"> <thead> <tr> <th>Criterio</th> <th>si/no</th> </tr> </thead> <tbody> <tr><td>1. xxxxxxxx? A B C</td><td>xxxxxxx</td></tr> <tr><td>2. xxxxxxxx? A B C</td><td>xxxxxxx</td></tr> <tr><td>3. xxxxxxxx? A B C</td><td>xxxxxxx</td></tr> <tr><td>4. xxxxxxxx? A B C</td><td>xxxxxxx</td></tr> <tr><td>5. xxxxxxxx? A B C</td><td>xxxxxxx</td></tr> </tbody> </table>	Criterio	si/no	1. xxxxxxxx? A B C	xxxxxxx	2. xxxxxxxx? A B C	xxxxxxx	3. xxxxxxxx? A B C	xxxxxxx	4. xxxxxxxx? A B C	xxxxxxx	5. xxxxxxxx? A B C	xxxxxxx	<table border="1"> <thead> <tr> <th>Criterio</th> <th>Núm. incumplimientos (frecuencia absoluta)</th> <th>% frecuencia (relativa)</th> <th>Frecuencia acumulada</th> </tr> </thead> <tbody> <tr><td>5. Consumo de alcohol</td><td>54</td><td>15.8</td><td>15.8</td></tr> <tr><td>12. Fondo de ojo</td><td>51</td><td>14.9</td><td>30.7</td></tr> <tr><td>11. Edemas</td><td>45</td><td>13.2</td><td>43.9</td></tr> <tr><td>4. Consumo de tabaco</td><td>42</td><td>12.3</td><td>56.1</td></tr> <tr><td>8. Exploración abdominal</td><td>36</td><td>10.5</td><td>66.7</td></tr> <tr><td>9. Pulsos periféricos</td><td>30</td><td>8.8</td><td>75.4</td></tr> <tr><td>10. Soplos carotídeos</td><td>27</td><td>7.9</td><td>83.3</td></tr> <tr><td>1. Diagnóstico correcto</td><td>15</td><td>4.4</td><td>87.7</td></tr> <tr><td>7. Auscultación cardiaca</td><td>15</td><td>4.4</td><td>92.1</td></tr> <tr><td>2. Antecedentes familiares</td><td>12</td><td>3.5</td><td>95.6</td></tr> <tr><td>6. Peso y talla y/o IMC</td><td>9</td><td>2.6</td><td>98.2</td></tr> <tr><td>3. Antecedentes personales</td><td>6</td><td>1.8</td><td>100.0</td></tr> <tr><td>Total</td><td>342</td><td>100</td><td></td></tr> </tbody> </table>  <p>Presentación y análisis de datos</p>	Criterio	Núm. incumplimientos (frecuencia absoluta)	% frecuencia (relativa)	Frecuencia acumulada	5. Consumo de alcohol	54	15.8	15.8	12. Fondo de ojo	51	14.9	30.7	11. Edemas	45	13.2	43.9	4. Consumo de tabaco	42	12.3	56.1	8. Exploración abdominal	36	10.5	66.7	9. Pulsos periféricos	30	8.8	75.4	10. Soplos carotídeos	27	7.9	83.3	1. Diagnóstico correcto	15	4.4	87.7	7. Auscultación cardiaca	15	4.4	92.1	2. Antecedentes familiares	12	3.5	95.6	6. Peso y talla y/o IMC	9	2.6	98.2	3. Antecedentes personales	6	1.8	100.0	Total	342	100	
Criterio	si/no																																																																						
1. xxxxxxxx? A B C	xxxxxxx																																																																						
2. xxxxxxxx? A B C	xxxxxxx																																																																						
3. xxxxxxxx? A B C	xxxxxxx																																																																						
4. xxxxxxxx? A B C	xxxxxxx																																																																						
5. xxxxxxxx? A B C	xxxxxxx																																																																						
Criterio	Núm. incumplimientos (frecuencia absoluta)	% frecuencia (relativa)	Frecuencia acumulada																																																																				
5. Consumo de alcohol	54	15.8	15.8																																																																				
12. Fondo de ojo	51	14.9	30.7																																																																				
11. Edemas	45	13.2	43.9																																																																				
4. Consumo de tabaco	42	12.3	56.1																																																																				
8. Exploración abdominal	36	10.5	66.7																																																																				
9. Pulsos periféricos	30	8.8	75.4																																																																				
10. Soplos carotídeos	27	7.9	83.3																																																																				
1. Diagnóstico correcto	15	4.4	87.7																																																																				
7. Auscultación cardiaca	15	4.4	92.1																																																																				
2. Antecedentes familiares	12	3.5	95.6																																																																				
6. Peso y talla y/o IMC	9	2.6	98.2																																																																				
3. Antecedentes personales	6	1.8	100.0																																																																				
Total	342	100																																																																					
<p>Causas principales</p> <p>1. _____</p> <p>2. _____</p>	<p>Solución</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>Resultados</p>  <table border="1"> <thead> <tr> <th></th> <th>Antes</th> <th>Después</th> <th>Manteniendo lo alcanzado</th> </tr> </thead> <tbody> <tr> <td>%</td> <td>___</td> <td>___</td> <td>___</td> </tr> <tr> <td>%</td> <td>___</td> <td>___</td> <td>___</td> </tr> <tr> <td>Fecha</td> <td>___</td> <td>___</td> <td>___</td> </tr> </tbody> </table>			Antes	Después	Manteniendo lo alcanzado	%	___	___	___	%	___	___	___	Fecha	___	___	___																																																				
	Antes	Después	Manteniendo lo alcanzado																																																																				
%	___	___	___																																																																				
%	___	___	___																																																																				
Fecha	___	___	___																																																																				

Figura 9.4. Narración de EMC

siones sobre el proyecto en las reuniones del grupo, e incluso de manera informal con todos aquellos que deben estar implicados en cambiar a mejor la manera de hacer las cosas.

Hay que tener presente que un ciclo de mejora no es un trabajo de investigación, y la comprobación de la mejora tras la intervención

no tiene por qué ser más pura o científica si se deja que la intervención actúe sola, sin sesgos. Si buscamos e identificamos alguna manera de favorecer la mejora no previsto o no explicitado en la estrategia diseñada, no hay que tener ningún reparo en ponerlo en práctica si funciona, es decir si conseguimos finalmente mejorar.

BIBLIOGRAFÍA

- Goldfield N, Pine M, Pine J. Techniques for analyzing processes and managing team decisions. In: Measuring and managing health care quality. Aspen, 1995. Vol I, Cap. 5.
- Leebow W, Ersoz CJ. Tools for targeting improvements. In: The health care manager's guide to continuous quality improvement. Chicago: AHA, 1991. Cap. 11.

Anexo 9.1

Indicaciones para la aplicación práctica de los contenidos de este capítulo

Una vez descrita la intervención o intervenciones a implantar (capítulo 8), se debe plasmar de forma gráfica un método de seguimiento, de forma que se pueda hacer control explícito de su implementación. La herramienta más común es un diagrama de Gantt. En su forma más completa figuran las actividades previstas, el tiempo que precisarán para ser implementadas, la secuencia en la que deben ser implementadas, y los responsables de llevarlas a cabo. Siguiendo el modelo de la figura 9.2 de este capítulo. Los pasos a seguir son los siguientes:

1. Establecer un cronograma explícito para la implementación de las actividades contenidas en el diagrama de afinidades del trabajo práctico anterior, plasmándolo en un diagrama de Gantt.
2. Utilizar el modelo propuesto en la figura 9.2 de este capítulo.
3. Procurar ser realista, de forma que se pueda tener una idea lo más aproximada posible del tiempo que llevará completar la intervención diseñada.
4. Considerar la conveniencia de construir un diagrama de Gantt por cada línea estratégica o grupo de acciones del diagrama de afinidades.

Una alternativa complementaria es utilizar un cartel narrativo del ciclo de mejora, preferentemente iniciado al principio del proyecto, e insertar en él el diagrama de Gantt.

Capítulo 10

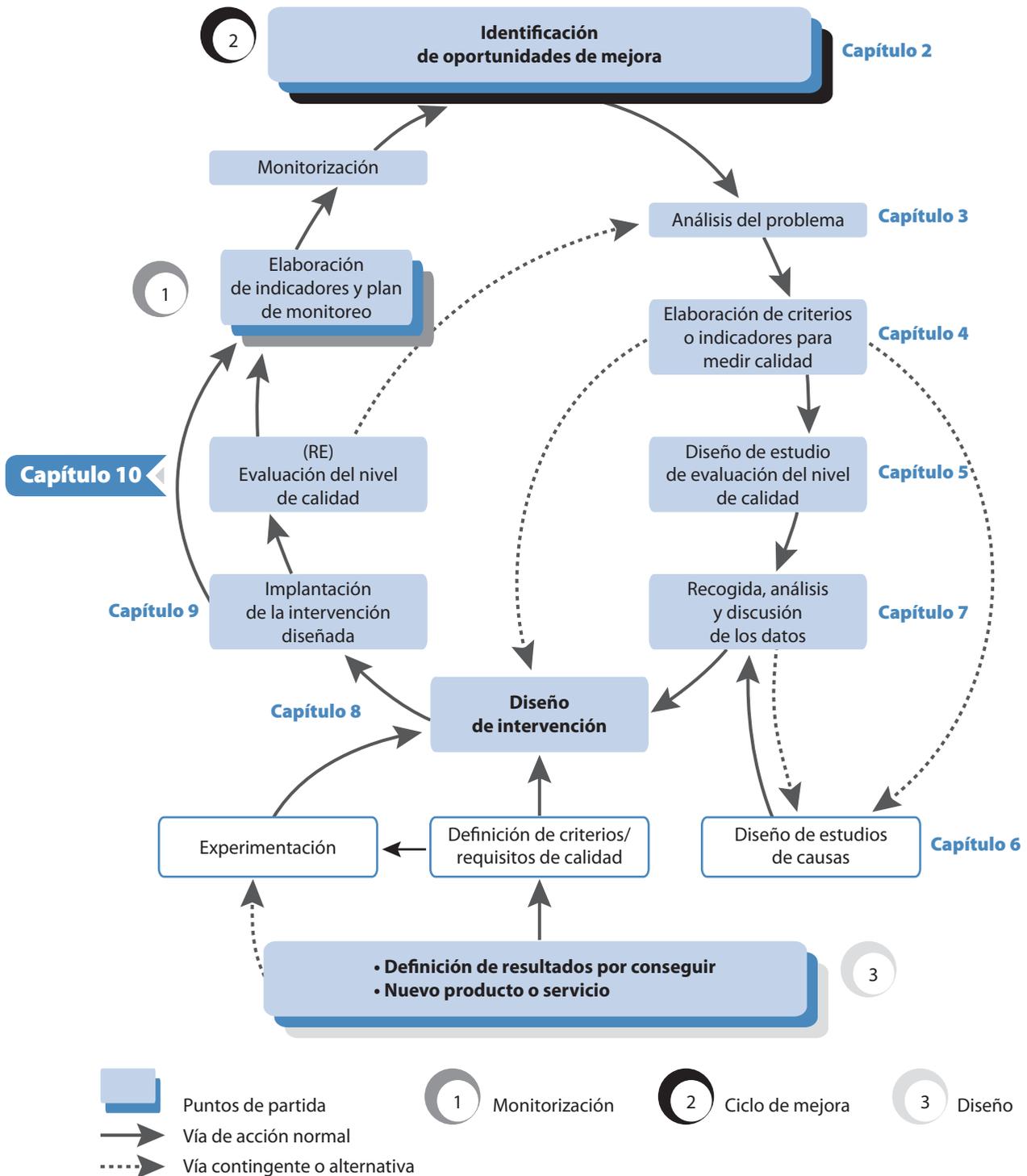
Cierre del ciclo de mejora: reevaluación, análisis y presentación de resultados comparativos de dos evaluaciones

Contenido general y objetivos específicos

1. Contextualización
2. Diseño de la reevaluación. Componentes y diferencias con el diseño de la primera evaluación
3. Presentación y análisis de los datos: estimación del nivel de calidad y mejora conseguida tras la intervención para mejorar.
4. ¿Hemos mejorado? Significación estadística de la diferencia entre las dos evaluaciones.
5. Representación gráfica de las diferencias en los niveles de calidad entre dos evaluaciones.
6. Análisis gráfico de la mejora y defectos por mejorar: diagrama de pareto antes-después.
7. ¿Qué hacemos a continuación? Cursos de acción tras la reevaluación

Anexo 10.1. Indicaciones para la aplicación práctica de los contenidos de este capítulo

Actividades para la mejora continua, pasos del ciclo de mejora y su correspondencia con los capítulos de este manual



Contenido general

Una vez que se haya dado tiempo a la intervención para surtir efecto, es precisa una nueva medición de la calidad para documentar la mejora. En este capítulo se detalla la metodología necesaria para cumplir con este objetivo. Se parte del hecho de que en la reevaluación lo que pretendemos no es solamente medir el nuevo nivel de calidad sino también comprobar la hipótesis de mejora, para ello tenemos que asegurarnos de que los datos a obtener sean comparables con la situación de partida, procurando que el diseño del estudio la reevaluación no introduzca modificaciones que impidan la comparabilidad. También se necesitan aplicar las pruebas estadísticas pertinentes y realizar un adecuado análisis gráfico de los datos.

Objetivos específicos

En este capítulo se describen los métodos para:

1. Diseñar la recogida de datos utilizando un esquema similar al empleado en la primera evaluación o evaluación basal de la calidad (capítulo 5).
2. Realizar la recogida de los datos en relación al nivel de calidad en esta evaluación, de forma similar a lo realizado con los datos de la primera evaluación. Calcular la mejora absoluta y relativa de cada criterio o requisito de calidad.
3. Aplicar correctamente diversas pruebas de significación estadística para el contraste de la hipótesis de mejora.
4. Interpretar los resultados de estas pruebas para responder a la pregunta ¿hemos mejorado?
5. Presentar gráficamente los resultados comparativos del nivel de calidad de las dos evaluaciones con un gráfico estrella o radar.
6. Construir un gráfico de Pareto antes-después.
7. Interpretar los resultados del ciclo de mejora efectuado, utilizando un gráfico de Pareto antes-después que indique la mejora conseguida y priorice las siguientes actividades a realizar.

Cierre del ciclo de mejora: reevaluación, análisis y presentación de resultados comparativos de dos evaluaciones

*“La vida sólo se comprende mirando hacia atrás pero se debe vivir hacia adelante”
Kierkegaard*

1. Contextualización

Después de haber puesto en marcha las medidas acordadas para mejorar el nivel de calidad, llega el momento, tras un cierto tiempo, de comprobar si efectivamente hemos logrado mejorar, y decidir en consecuencia qué debemos hacer a continuación en nuestro camino de mejora continua.

Documentar la mejora y decidir qué hacer a continuación entraña los siguientes pasos: (i) diseñar la (re)evaluación de forma que los datos sean comparables con la primera evaluación (basal); (ii) analizar los datos correctamente para determinar cuál es el nivel de calidad alcanzado y responder a la pregunta ¿hemos mejorado?; y (iii) analizar los datos para responder a la pregunta ¿qué hacemos a continuación para seguir mejorando? A lo largo de este proceso se han de ir tomando decisiones metodológicas concretas que son la base de este capítulo. En primer lugar se contemplan las diferencias posibles y no deseables en los diseños de la primera y segunda evaluación para asegurar la comparabilidad, y a continuación se plantea cómo realizar el análisis comparativo de los datos obtenidos con los que obtuvimos en la evaluación basal.

2. Diseño de la reevaluación. Componentes y diferencias con el diseño de la primera evaluación

El diseño metodológico de la evaluación para documentar la mejora conseguida no tiene por qué tener ninguna diferencia con el que utilizamos para la primera evaluación. Sus componentes son idénticos (se reproducen en

el cuadro 10.1) y responden a la misma descripción que se describe en el capítulo 5, que es conveniente revisar a este respecto.

Sin embargo, tras la experiencia de la primera evaluación se plantean a veces algunas dudas en relación al mantenimiento de exactamente los mismos criterios, las peculiaridades de las unidades de estudio que pueden condicionar el marco muestral para la segunda evaluación, y la conveniencia de utilizar tamaños de muestra diferentes. Vamos a revisar las causas más frecuentes por las que se pueden plantear estas cuestiones y la forma más razonable de afrontarlas.

2.1. Cambios en los criterios o indicadores de la calidad en la segunda evaluación

Para efectos de comparabilidad con la primera evaluación no es deseable en absoluto cambiar

En este capítulo se abordan los aspectos metodológicos relacionados con el diseño de la segunda evaluación para la comparación con la situación basal.

Cuadro 10.1 Componentes de una evaluación de la calidad	
1.	Criterios para evaluar la calidad
2.	Dimensión estudiada
3.	Tipos de datos
4.	Unidades de estudio
5.	Fuentes de datos
6.	Identificación y muestreo de la unidades de estudio
	● Marco muestral
	● Número de casos
	● Métodos de muestreo
7.	Tipo de evaluación
	● Según relación temporal
	● Según quién tomó la iniciativa
	● Según quién obtiene los datos

El diseño de la reevaluación es conveniente que sea en general idéntico al de la primera evaluación. Sin embargo la experiencia de la primera evaluación puede habernos sugerido introducir cambios que hay que considerar con precaución.

Para que se puedan comparar, los criterios han de ser definidos y medidos de forma semejante en las dos evaluaciones. Los cambios de definición los convierte en nuevos criterios para los que se realiza una primera evaluación.

Aunque pueden aceptarse diferencias en el marco muestral (universo para la extracción de casos), no debe cambiarse la definición de la unidad de estudio entre la primera y la segunda evaluación.

Si hemos utilizado criterios de comparabilidad problemática (cumplimientos en periodos de tiempo únicos o muy largos) junto a otros más sencillos, siempre queda la opción de realizar su reevaluación por separado.

la definición de criterios a evaluar. Cuando se plantea esta duda se debe casi siempre a defectos en el diseño de los criterios evaluados en la primera evaluación. Frecuentemente, estos defectos se deben a una falta de concreción que los hace difíciles de valorar; por ejemplo si un criterio incluye varios subapartados, tras evaluarlo puede evidenciarse que deberían haber sido considerados aisladamente, o incluso que algunos ítems faltan o sobran. Son ejemplos frecuentes los criterios sobre anamnesis de antecedentes personales, exploración física básica, etc., que suelen incluir en realidad varios subcriterios y podemos decidir, después de haberlos medido, que no es muy informativo medirlos conjuntamente como un solo criterio. Otras veces lo que ocurre es que nos damos cuenta que hemos incluido criterios irrelevantes o que se cumplen siempre, mientras otros más importantes y/o problemáticos no han sido medidos; estamos tentados entonces de incluir nuevos criterios para la segunda evaluación, e incluso eliminar aquellos que se nos han revelado como irrelevantes.

El problema común a todos estos cambios, que puede estar muy justificado hacer, es el de la comparabilidad entre las dos evaluaciones: cualquier nuevo criterio o nueva forma de entender (o describir) un criterio lo convierte de hecho en una primera evaluación y la comparación con la anterior evaluación no es posible ni deseable. La comparación se debe efectuar siempre entre criterios que hayan sido descritos, entendidos y medidos de forma semejante en las dos evaluaciones.

2.2. Diferencias en la definición de las unidades de estudio y en el marco muestral

Al igual que para los criterios, no es deseable para la reevaluación cambio alguno en la definición de las unidades de estudio ni, en menor grado y con excepciones, en las características del marco muestral.

Sin embargo, la primera vez que se evalúa la calidad de la atención a un determinado problema de salud son frecuentes los marcos muestrales definidos en términos de *todos los casos* existentes en el centro y, si se evalúan criterios a los que corresponde como unidad de estudio un periodo de tiempo único en todo

el proceso de atención (como por ejemplo el diagnóstico), o muy extenso (por ejemplo supongamos que las exploraciones electrocardiográficas en los pacientes hipertensos son establecidas cada tres años como criterio de buen seguimiento), pueden surgir dudas sobre cómo definir los casos y/o el marco muestral en la reevaluación, dada la dificultad de que se puedan reunir un número suficiente de casos para reevaluar los criterios con periodos de tiempo y/o marcos muestrales como los definidos en la primera evaluación. ¿Qué es lo más conveniente entonces para la re-evaluación?, ¿una muestra de todos los casos?, ¿Sólo los casos nuevos que se hayan producido tras la primera evaluación?

Criterios con periodo de tiempo único en el proceso por evaluar. ¿Todos los casos o sólo casos nuevos?

Tomemos por ejemplo el criterio sobre diagnóstico de la hipertensión. Está claro que si definimos la unidad de estudio como las 2 semanas a 3 meses en las que se deben hacer las mediciones para clasificar al paciente como hipertenso (en ese "segmento" irrepetible del proceso evaluamos si se procedió correctamente o no), debe mantenerse la misma definición para la reevaluación, de forma que sólo deberíamos considerar en ella los *casos nuevos* diagnosticados desde que pusimos en marcha la intervención para mejorar; sin embargo, el marco temporal para la muestra de casos sería muy diferente, si hemos utilizado como universo para la primera evaluación todos los hipertensos diagnosticados hasta el momento de la evaluación; esta diferencia en el marco muestral es aceptable aunque hubiese sido quizá mejor elegir como marco muestral para la primera evaluación los diagnosticados en un cierto periodo de tiempo semejante al que dejaríamos pasar entre intervención y reevaluación. ¿Qué ocurriría si decidimos mantener el mismo marco muestral, y realizar de nuevo una muestra extraída de todo el universo de hipertensos? Mantener el mismo marco muestral (todos los hipertensos diagnosticados hasta el momento de la 2ª evaluación) es, efectivamente, una opción posible, pero en este caso el efecto de la intervención puede quedar diluido, aparte de

que tendríamos que decidir qué hacer con los diagnosticados incorrectamente antes de la primera evaluación: excluirlos de la segunda evaluación o iniciar de nuevo el proceso diagnóstico con ellos e incluirlos en el marco muestral de la segunda evaluación si han visitado el centro entre las dos evaluaciones; en esta segunda opción, estamos asimismo cambiando el criterio en cuanto al periodo de tiempo contemplado en la primera evaluación porque consideramos la “segunda oportunidad” no contemplada en la primera evaluación. Obviamente, con estos cambios estamos introduciendo “ruido” en la comparación. En el primer caso (si consideramos de nuevo todos los hipertensos en vez de solo los nuevos diagnósticos) conducente a diluir el posible efecto de la intervención; y en el segundo (considerando también como nuevos diagnósticos el “re-diagnóstico” de los previamente mal diagnosticados) aumentándolo, a costa de redefinir el criterio hasta el punto que se podría considerar no comparable con el de la primera evaluación. Por todo ello, hay que considerar muy cuidadosamente cualquier cambio en la definición de la unidad de estudio y marco muestral de la segunda evaluación, reflexionando sobre los posibles sesgos que estos cambios puedan introducir en la comparabilidad e interpretación de los datos.

A efectos comparativos es siempre más importante *mantener la definición de la unidad de estudio*, que la definición del marco muestral o universo para la extracción de los casos a evaluar, teniendo a veces que esperar más tiempo para reunir casos suficientes para evaluar. Si introducimos cambios en los criterios o periodos de tiempo en el que el criterio debe cumplirse, se trata de estudios complementarios, probablemente necesarios y justificados, pero no directamente comparables con la primera evaluación.

Criterios con periodo de tiempo largo. ¿Sólo nuevos casos o incluimos los que no cumplan el criterio en la primera evaluación?

En el caso de criterios con periodo de tiempo muy largo a la hora de definir la unidad de estudio (como por ejemplo un ECG cada 3 años en el control del hipertenso), podríamos dudar en la reevaluación si incluir aquellos que ya cumplían el criterio en la primera evaluación

(un acercamiento del tipo “todos los casos” existentes, al igual que en la primera evaluación que acabamos de comentar) o evaluar sólo los casos que no la cumplieran y que han visitado el centro desde que decidimos intervenir, más los que no se habían incluido en la primera evaluación por no llevar suficiente tiempo diagnosticados (menos de tres años) y ya lo llevan al reevaluar. En este caso, debe primar en la decisión el mantenimiento de la misma definición de unidad de estudio y revisar los últimos tres años de todos los que llevan al menos este tiempo diagnosticados en ambas evaluaciones, porque actuar correctamente (cumplir el criterio) implica valorar continuamente si se tiene o no se tiene hecho el ECG en los últimos tres años. Tanto la unidad de estudio como el marco muestral pueden y deben ser, en este caso, iguales en las dos evaluaciones. Pueden incluirse por tanto aquellos casos excluidos en la primera evaluación porque no cumplían aún con el periodo de tiempo necesario para ser evaluados.

Finalmente, en el caso de que hayamos evaluado criterios de fácil comparabilidad en la segunda evaluación junto a otros potencialmente problemáticos por sus peculiares unidades de tiempo (únicas o muy largas) a que se refieren, no es mala solución evaluarlos por separado y tras un tiempo diferente tras la intervención; es decir, esperar seis meses o hasta un año para algunos de ellos, mientras que se evaluarán antes aquellos en los que el efecto de la intervención pueda verse a más corto plazo.

Cuando los criterios se refieren a periodos de tiempo en el proceso de atención que no son únicos o muy largos, no suelen plantearse este tipo de dudas y problemas potenciales para la comparabilidad de las dos evaluaciones. Es el caso por ejemplo de todos los criterios que se refieren a problemas agudos o, en caso de pacientes crónicos, a cada visita, o en la última visita.

2.3. Diferencias en el tamaño de la muestra.

A diferencia de la definición de criterio y su correspondiente unidad de estudio, el tamaño de la muestra para la segunda evaluación *no tiene que ser necesariamente igual que el de la primera*. Es conveniente que sea manejable, se obtenga de forma aleatoria, si el universo es suficien-

El tamaño de la muestra puede ser diferente en la segunda evaluación, siempre que se respete la aleatoriedad de su extracción. Para detectar como significativas mejoras pequeñas (inferiores a 0,15 ó 0,1) necesitamos muestras más grandes, sobre todo si el cumplimiento en la 1ª evaluación fue en torno a 50%.

te grande como para aconsejar una muestra, y que tenga, como para la primera evaluación, un tamaño que permita estimar el nivel de calidad con una cierta precisión. Nada más.

En el capítulo 5 propusimos la cifra de 50-60 casos como un tamaño razonablemente pequeño que permitía inferencias de una precisión normalmente suficiente. Sin embargo al reevaluar no estamos interesados solamente en estimar el nivel de calidad: también queremos saber con una cierta seguridad si hemos mejorado o no. Esta seguridad nos la da, como veremos más adelante, el cálculo del *nivel de significación* de la diferencia encontrada entre las dos evaluaciones; nivel de significación en el que influye el tamaño de la muestra en *ambas evaluaciones*. Como norma general, muestras pequeñas van a dar como significativas diferencias grandes, mientras que hacen falta muestras más grandes para que diferencias (en nuestro caso mejoras) modestas aparezcan como significativas. Quiere esto decir que si queremos aumentar las posibilidades de detectar como significativas diferencias no muy grandes entre las dos evaluaciones, nos conviene aumentar el tamaño de la muestra en ambas o al menos en la segunda evaluación.

Para tener una idea de en qué medida necesitamos o no muestras más grandes para detectar según qué diferencias, en el cuadro 10.2 figura el tamaño de la muestra en ambas eva-

luaciones que sería necesario para que fuesen significativos diversos niveles de diferencia entre las dos evaluaciones, en función del valor encontrado en la primera evaluación. Tanto el concepto de significación estadística, como el fundamento del cuadro 10.2 serán explicados en otros apartados de este capítulo. Baste, por ahora resumir que una muestra de 60 casos va a ser suficiente para detectar de forma significativa diferencias que van desde un mínimo de 0.15, cuando la proporción de cumplimiento en la primera evaluación fue de en torno a 0.5 (50%), hasta un mínimo de 0.05 o 0.1 cuando las proporciones de cumplimiento en la primera evaluación son más extremas (0.05, 0.1 o 0.95). En general, diferencias entre estimaciones de las dos evaluaciones iguales o superiores a 0.15 van a ser evidenciadas como significativas con muestras de 60 casos, sea cual sea la proporción de cumplimiento en la 1ª evaluación, siendo posible detectar diferencias tanto más pequeñas cuando más alejado está del 50% el cumplimiento en la primera evaluación.

3. Presentación y análisis de los datos: estimación del nivel de calidad y mejora conseguida tras la intervención para mejorar

Una vez efectuada la nueva recogida de datos, nos encontramos ante una situación semejante

Cuadro 10.2

Tamaño de muestra (n) necesario en cada evaluación para detectar como significativas ($p \leq 0.05$, una cola) diversos niveles de diferencias entre dos evaluaciones (d), según la proporción de cumplimiento de la primera evaluación (p_1)

d	Proporción de cumplimiento en la primera evaluación (p_1)																		
	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
0.05	150	237	313	377	432	475	507	529	540	540	529	507	475	432	377	313	237	150	53
0.08	69	102	130	154	174	190	201	208	211	210	205	195	181	163	140	114	83	48	32
0.10	49	69	87	101	114	123	130	134	135	134	130	123	114	101	87	69	49	*	*
0.13	33	44	54	62	69	74	78	80	80	79	76	71	65	58	48	37	*	*	*
0.15	*	35	42	48	53	56	59	60	60	59	56	53	48	42	35	*	*	*	*
0.18	*	*	30	34	37	40	41	42	41	40	38	36	32	*	*	*	*	*	*
0.20	*	*	*	*	31	32	33	34	33	32	31	*	*	*	*	*	*	*	*
0.23	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0.25	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

(*): $n < 30$

d: diferencia entre las dos evaluaciones: $p_2 - p_1$

a la de la primera evaluación en cuanto a las estimaciones o inferencias que podemos hacer. Los métodos estadísticos en este sentido van a ser idénticos a los de la primera evaluación, tanto para realizar la estimación puntual como para el intervalo de confianza de los niveles de calidad alcanzados. Sin embargo, tendremos ahora el interés adicional de *comparar* ambas evaluaciones y ver si hemos logrado mejorar. Para ello, si hemos evaluado con muestras, tendremos que tener en cuenta, también para la comparación, los posibles errores en las estimaciones (puntual e intervalo), en ambas

evaluaciones. En consecuencia vamos a realizar una estimación puntual de la diferencia y *además* comprobar si esta diferencia puede ser atribuible o no a los errores esperables por haber medido con muestras; esta comprobación se hace en base a un cálculo de probabilidades y se expresa, en la jerga estadística, como *significación estadística* de la diferencia entre las dos evaluaciones. Vamos a revisar lo que entraña conceptual y operativamente cada uno de esos pasos, siguiendo el resumen que figura en el cuadro 10.3 y ejemplificando los cálculos con los datos del cuadro 10.4.

En la reevaluación estamos interesados en saber el nivel de calidad alcanzado pero también, y sobre todo, documentar la mejora conseguida; es decir: comparar adecuadamente los resultados de las dos evaluaciones.

Cuadro 10.3	
Análisis de los datos de una reevaluación. Resúmen de componentes y métodos	
Tipo de muestreo	Estimación puntual
<p>1. Estimación del nivel de calidad</p> <ul style="list-style-type: none"> - Estimación puntual: - Intervalo de confianza (95%): 	$P_2 = \frac{\text{nº de cumplimientos}}{\text{tamaño de la muestra}}$ $P_2 \pm 1.96 \sqrt{\frac{p_2(1-p_2)}{n}}$
<p>2. Estimación de la mejora conseguida</p> <ul style="list-style-type: none"> - Mejora absoluta: - Mejora relativa: 	$P_2 - P_1 =$ $\frac{P_2 - P_1}{1 - P_1}$
<p>3. Significación estadística de la mejora</p> <ul style="list-style-type: none"> - Calcular valor de z para la diferencia entre las dos evaluaciones (**): - Diferencia entre proporciones: - Diferencia entre medidas: - Ver significación en tablas de la distribución normal (una cola): 	$z = \frac{P_2 - P_1}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$ $z = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
<p>p: proporción de cumplimiento conjunta (primera y segunda evaluación juntas) = $\frac{\text{nº cumplimientos en la 1ª} + \text{nº cumplimientos en la 2ª}}{n_1 + n_2}$</p> <p>$\bar{x}_2$: media en la reevaluación; \bar{x}_1: media en la primera evaluación.</p> <p>s2: desviación estándar de los datos de la muestra en la reevaluación; s1: desviación estándar de los datos de la muestra en la primera evaluación.</p>	

Cuadro 10.4

Evaluación de la calidad del diagnóstico, anamnesis y exploración física inicial del paciente hipertenso antes y después de la intervención para mejorar. Niveles de cumplimiento de los criterios en una muestra aleatoria de 60 casos

Criterio*	1ª Evaluación		2ª Evaluación		
		%Cumplimiento (ic 95%)		%Cumplimiento (ic 95%)	
1.	Diagnóstico correcto	75.0	(±11.0)	95.0	(±5.5)
2.	Antecedentes familiares	80.0	(±10.1)	86.0	(±8.8)
3.	Antecedentes personales	90.0	(±7.6)	98.0	(±3.5)
4.	Consumo de tabaco	30.0	(±11.6)	70.0	(±11.6)
5.	Consumo de alcohol	10.0	(±7.6)	50.0	(±12.7)
6.	Peso y talla y/o IMC	85.0	(±9.0)	84.0	(±9.3)
7.	Auscultación cardiaca	75.0	(±11.0)	80.0	(±10.1)
8.	Exploración abdominal	40.0	(±12.4)	57.0	(±12.5)
9.	Pulsos periféricos	50.0	(±12.7)	60.0	(±12.4)
10.	Soplos carotídeos	55.0	(±12.6)	75.0	(±11.0)
11.	Edemas	25.0	(±11.0)	50.0	(±12.7)
12.	Fondo de ojo	15.0	(±9.0)	25.0	(±11.0)

Las estimaciones puntual y de intervalo del nivel de calidad se realizan de forma idéntica a las de la primera evaluación.

3.1. Estimación del nivel de calidad

El objetivo y los procedimientos son idénticos a los que vimos en el capítulo 7. Los datos de nuestra segunda evaluación nos sirven de base para una estimación puntual del nivel de cumplimiento de los criterios, y la estimación de un intervalo de valores (intervalo de confianza) entre los cuales tenemos un nivel considerable de certeza (confianza, normalmente establecida en 95%) de que se encuentra el valor real.

Las fórmulas más usuales son las que figuran en el cuadro 10.3, pero existe una explicación más extensa en el capítulo 7, al cual debemos remitirnos en caso de duda en este punto. El cuadro 10.4 contiene ambas estimaciones, puntual y de intervalo, para los criterios considerados en dos evaluaciones.

3.2. Estimación de la mejora conseguida

La forma más directa e intuitiva de saber si hemos mejorado es simplemente ver la diferencia entre los niveles de cumplimiento de la segunda y la primera evaluación. Esta diferencia es lo que podemos denominar estimación de la mejora absoluta para cada criterio considerado. Por ejemplo, según los datos del cuadro

10.4 la mejora absoluta para los criterios 1, 3, 5 y 10 es de 20, 8, 40 y 20 puntos respectivamente (95-75, 98-90, 50-10 y 75-55). De igual manera podemos calcular la mejora absoluta para cada uno de los criterios del cuadro 10.4 y veremos que en todos ellos es positiva, excepto para el criterio 6. Sin hacer ningún otro tipo de consideraciones, parece que se ha logrado mejorar de forma consistente en prácticamente todos los criterios.

Pero volvamos al ejemplo de los cuatro criterios (1, 3, 5 y 10) que hemos seleccionado para ejemplificar el cálculo de la mejora absoluta; podemos observar que, sólo con las cifras de mejora absoluta, parece que los criterios 1 y 10 han mejorado en igual medida (20 puntos cada uno), el criterio 5 ha mejorado el doble que estos dos, y el criterio 3 ha mejorado comparativamente bastante menos que los anteriores. Sin embargo, estas apreciaciones no tienen en cuenta que el punto de partida, y por tanto las posibilidades o espacio para mejorar, son muy diferentes para cada uno de estos criterios. No es lo mismo normalmente subir de 75% a 95%, como ocurre con el criterio 1, que de 55% a 75% como ocurre con el criterio 10, aunque en ambos casos la mejora absoluta sea de 20 puntos; de igual manera, los 40 puntos de mejora del

criterio 5 son posibles porque se parte de unos niveles de cumplimiento muy bajos, que no se dan en el criterio 3, con unos niveles de 90% de cumplimiento en la primera evaluación que sólo permiten mejorar como máximo 10 puntos (de 90 a 100%). Con el objeto de considerar explícitamente esos matices y poder comparar mejor el esfuerzo de mejora entre criterios y/o entre centros para los mismos criterios, es conveniente calcular lo que podemos llamar *mejora relativa*, que consiste en relativizar la mejora absoluta en relación al espacio de mejora posible total que existía en la primera evaluación; este *espacio de mejora posible* es 1 (o 100%, máximo cumplimiento posible) menos el nivel de cumplimiento en la primera evaluación (en proporción o porcentaje, según sea 1 o 100 como hayamos caracterizado el nivel máximo de cumplimiento). La fórmula sería pues:

$$\frac{\text{Mejora absoluta}}{\text{espacio posible de mejora}} = \frac{P_2 - P_1}{1 - p_1}$$

siendo p_2 el cumplimiento en la segunda evaluación y p_1 el cumplimiento que encontramos en la primera evaluación. Para los cuatro criterios que hemos elegido como ejemplo, la *mejora relativa* sería la siguiente:

- Criterio 1 (Diagnostico):

$$\frac{0.95 - 0.75}{1 - 0.75} = \frac{0.2}{0.25} = 0.8 \text{ o } 80\%$$

- Criterio 3 (Antecedentes personales):

$$\frac{0.98 - 0.90}{1 - 0.90} = \frac{0.08}{0.1} = 0.8 \text{ o } 80\%$$

- Criterio 5 (Consumo alcohol):

$$\frac{0.50 - 0.10}{1 - 0.10} = \frac{0.4}{0.9} = 0.44 \text{ o } 44\%$$

- Criterio 10 (Soplos carotídeos):

$$\frac{0.75 - 0.55}{1 - 0.55} = \frac{0.20}{0.45} = 0.44 \text{ o } 44\%$$

Con este nuevo cálculo vemos que en términos *relativos* los criterios que más han mejorado con el 1 y el 3, en los que se ha logrado aprovechar el 80% de la mejora posible, mientras que los mismos 20 puntos de mejora absoluta

del criterio 10 sólo suponen el 44% de la mejora posible. Es evidente que esta es una valoración más ajustada del esfuerzo realizado cuando queramos hacer comparaciones.

En general, es conveniente calcular tanto la mejora absoluta como la mejora relativa. No obstante, tanto la mejora absoluta como la relativa las hemos calculado en base a las estimaciones puntuales de los niveles de cumplimiento que sabemos están sujetos a los errores propios de haber utilizado una muestra. Ello quiere decir que, sin hacer algunos cálculos adicionales, no sabremos hasta qué punto estamos seguros que la mejora es real, o si puede responder a los posibles valores que podemos encontrar por la variabilidad esperable al utilizar muestras, aunque no exista diferencia en la realidad entre las dos evaluaciones. Es necesario, pues, ser capaces de afirmar hasta qué punto la diferencia encontrada es compatible o no con la inexistencia de mejora en la realidad: hay que calcular la *significación estadística* de la diferencia encontrada entre las dos evaluaciones.

4. ¿Hemos mejorado? Significación estadística de la diferencia entre las dos evaluaciones

Con la estimación puntual de la diferencia entre las dos evaluaciones ocurre lo mismo que veíamos al dar la estimación puntual del nivel de calidad: no sabemos con sólo dar el valor de la estimación hasta qué punto puede responder o no al valor real; la cuestión de base es si las dos estimaciones puntuales, es decir los dos niveles de cumplimiento de la primera y la segunda evaluación, pueden responder a valores probables en una muestra de un mismo valor real, o, lo que es lo mismo, si la diferencia observada con las dos muestras puede ser explicada por los errores esperables por el muestreo. Esta incógnita puede ser, desde luego, despejada con la ayuda de los cálculos estadísticos oportunos.

Hay dos maneras de aclararlo: (i) estimando un intervalo de valores de la diferencia; o (ii) con un test estadístico para comprobar la probabilidad de que la diferencia real entre las dos evaluaciones sea cero.

La primera opción ya nos es *familiar*: estimamos el intervalo de confianza *de la diferencia*,

Para estimar la mejora conseguida es conveniente calcular no solo la diferencia absoluta entre las dos evaluaciones (mejora absoluta) sino también relativizarla al espacio de mejora posible que teníamos en la primera evaluación (mejora relativa).

Hay que realizar algunos cálculos estadísticos para poder concluir si la diferencia encontrada es real o puede ser explicada por los errores del muestreo.

Una forma de ver si la diferencia es real es calculando el intervalo de confianza de la diferencia. Si este intervalo no contiene el valor cero, podemos concluir, con la confianza fijada, que la diferencia es real y no debida a los errores de muestreo.

y si éste contiene el valor cero, no podremos afirmar que la diferencia es real. Recuérdese que el intervalo de confianza es un rango de valores entre los cuales sabemos con una cierta confianza (probabilidad de que sea cierto) que está el valor real del parámetro estimado. En este caso, el parámetro estimado con las muestras de las dos evaluaciones vamos a considerar que es la diferencia entre ellas, en vez del nivel de cumplimiento que es para lo que hemos utilizado la estimación de un intervalo de valores hasta ahora. La fórmula y razonamiento para el intervalo de confianza de la diferencia es en todo semejante a la que vimos en el capítulo 7, para el intervalo de confianza del nivel de cumplimiento. Responde a la fórmula general:

Parámetro estimado $\pm z \cdot$ Desviación estándar del parámetro estimado

El parámetro estimado es ahora la diferencia entre las dos evaluaciones ($p_2 - p_1$); z es 1.96, para una confianza de 95%; y el error estándar (o desviación estándar del parámetro estimado) es la suma del error estándar de las dos evaluaciones, de modo que el intervalo va a ser:

$$(p_2 - p_1) \pm 1.96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Si lo aplicamos por ejemplo a la diferencia que hemos encontrado para el criterio 1, en donde $p_2=0.95$ y $p_1=0.75$, con muestras de 60 casos en las dos evaluaciones, el intervalo de confianza de 95% para la diferencia sería:

$$(0.95-0.75) \pm 1.96 \sqrt{\frac{(0.75)(0.25)}{60} + \frac{(0.95)(0.05)}{60}} = 0.20 \pm 0.12$$

es decir, de 0.08 a 0.32; todos ellos valores por encima de cero con lo que podemos concluir, con una confianza de 95%, que la diferencia es real y además que es como mínimo de 8 puntos.

El segundo proceder, la utilización de un *test estadístico* para aceptar o rechazar la hipótesis de que no hay diferencia, nos puede resultar en general más conveniente por ser más adaptable a nuestra situación, en el sentido de que va a ser posible comprobar de forma aislada si la

diferencia es *a mejor*, que es exactamente lo que nos interesa, mientras que el intervalo de confianza incluye las posibles diferencias en los dos sentidos (a mejor y a peor), lo cual conlleva unas exigencias mayores en cuanto a detectar diferencias significativas. Para entender toda esta argumentación a favor de la utilización de un test estadístico, hay que introducir y aclarar dos nuevos conceptos como son la *significación estadística* y la formulación de las llamadas *hipótesis nula y alternativa* que son las que vamos a poner a prueba con el test estadístico.

4.1. ¿Qué es la significación estadística?

La *significación estadística* es un concepto que se utiliza normalmente en la comprobación de hipótesis y responde a la *probabilidad* de que esa hipótesis que se comprueba sea cierta, de manera que si esa probabilidad es muy baja, nos arriesgamos a rechazar que sea cierta y aceptamos que existe una situación diferente. En la jerga estadística a la hipótesis de partida que se acepta o rechaza con el test se le llama *hipótesis nula*, y lo que aceptamos cuando ésta se rechaza debido a que la probabilidad de que sea cierta es muy baja, se conoce como *hipótesis alternativa*. En nuestro caso la hipótesis nula sería que la diferencia entre las dos evaluaciones es en realidad cero, es decir, que no ha habido mejora y la diferencia estimada es explicable por los errores del muestreo; mientras que la hipótesis alternativa es que sí ha habido mejora y la diferencia encontrada responde a la existencia real de diferencia entre las dos evaluaciones y no al azar de los errores del muestreo. En resumen, lo que hacemos para aplicar un test estadístico para aceptar o rechazar una determinada hipótesis (o suposición de que algo es de una determinada manera) es lo siguiente:

1. Formulamos la hipótesis *nula*, o hipótesis de partida que normalmente ha de implicar la igualdad o inexistencia de cambios o diferencias. En nuestro caso, la hipótesis nula va a ser que no ha habido diferencia entre las dos evaluaciones.
2. Formulamos la hipótesis alternativa, que será la que vamos a aceptar si rechazamos la hipótesis nula. Nuestra hipótesis alternativa va a ser que sí que hay diferencia *a mejor* entre las dos evaluaciones

Para concluir si hemos mejorado más que calcular el intervalo de confianza de la diferencia, conviene utilizar un test estadístico apropiado.

(vamos a ver enseguida por qué subrayamos a *mejor*).

3. Aplicamos un test estadístico apropiado que nos va a decir la probabilidad de que la hipótesis nula sea cierta. Esa probabilidad es lo que vamos a llamar *significación estadística* de nuestra decisión.
4. En función de esa probabilidad vamos a aceptar o rechazar la hipótesis nula (y consiguientemente rechazar o aceptar la hipótesis alternativa).

Pero ¡jojo!: lo que puede resultar confuso es que decimos que el test estadístico es muy significativo cuando la probabilidad de que la hipótesis nula sea cierta es muy pequeña; es decir, más significación estadística cuanto menor es la probabilidad de que la hipótesis nula sea cierta.

En realidad, esta probabilidad es esencialmente el *riesgo de equivocarnos* que corremos al rechazar la hipótesis nula, puesto que hemos dicho que es la probabilidad de que sea cierta: es un riesgo que se conoce también como error a o Tipo I de la decisión tomada, y se anota habitualmente como el valor p (de probabilidad) del test estadístico. Como probabilidad de error que es, nos interesa adoptar una determinada como límite de decisión que no sea muy alta pero que nos dé una cierta seguridad en la decisión (recuérdese que p alta quiere decir significación *baja*); lo habitual es aceptar que sea como máximo de un 5%, de forma que si el valor de p que encontramos en el test es $\leq 5\%$ (en forma de proporción sería ≤ 0.05) vamos a rechazar la hipótesis nula, aceptaremos la hipótesis alternativa, y vamos a decir que el test es *significativo*, con un *nivel de significación igual a p* .

Una vez entendido el concepto de significación estadística y su aplicación a la aceptación o rechazo de las hipótesis nula (no *diferencia*) y alternativa (mejora significativa), podemos aplicarlo a nuestro caso. Sólo nos falta seleccionar el test estadístico apropiado y entender qué quiere decir y cómo se aplica el que la significación estadística sea de *una cola*.

4.2. Selección del test estadístico apropiado

Existen varios test estadísticos que podrían aplicarse para comprobar si hay o no diferencia real entre las dos evaluaciones. Estos tests

se diferencian entre sí básicamente por la distribución de probabilidades que utilizan para ver la significación (el valor de p), lo que conlleva unas determinadas condiciones de uso, y fórmula de cálculo. Por ejemplo, para comparar dos proporciones como es el caso de los niveles de cumplimiento de un criterio en las dos evaluaciones, podríamos utilizar una χ^2 , el test exacto de Fisher o el valor de z que corresponde a la diferencia y errores estándar de las dos muestras. De todos ellos, seleccionamos el test del valor de z por la facilidad de cálculo, junto a la posibilidad de ver la significación de una cola, algo cuya importancia veremos enseguida. La aplicación de este test se basa en el hecho de que los diversos valores posibles de la diferencia entre las dos evaluaciones que podemos encontrar por muestreo siguen en cuanto a su probabilidad de aparición una distribución normal; para ver la probabilidad de cada uno de los valores si la diferencia real es cero, se calcula el valor de z (valor estadístico estándar de la distribución normal), que corresponde a la diferencia encontrada.

La fórmula de cálculo es (tal como figura en el cuadro 10.3):

$$z = \frac{p_2 - p_1}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

donde p_2 es la proporción de cumplimiento en la segunda evaluación, p_1 es la proporción de cumplimiento en la primera evaluación, n_1 y n_2 el tamaño de la muestra en la primera y segunda evaluación, y p la proporción de cumplimiento conjunta de las dos evaluaciones (suma de cumplimientos de ambas evaluaciones dividida por la suma de ambas muestras). Una vez realizado el cálculo, miramos en las tablas de la distribución normal cuál es la probabilidad (significación) que corresponde al valor de z encontrado.

Como es evidente al estar $p_2 - p_1$ en el numerador, si no hay diferencia entre las dos evaluaciones el valor de z es cero, y obtendremos valores de z tanto mayores cuanto mayor sea la diferencia encontrada. La cuestión es ¿cuánto puede llegar a ser de grande el valor de z y seguir siendo compatible con una situación

La significación estadística para la decisión que se toma con la aplicación de un test estadístico es la probabilidad de que la hipótesis nula sea cierta. Se conoce también como error, error tipo I y valor p .

De entre los tests estadísticos que se pueden utilizar, seleccionamos el cálculo del valor z (valor estadístico en la distribución normal) para la diferencia encontrada.

real de no diferencia, es decir con un valor real de diferencia igual a cero?, dicho de otra forma, si para rechazar que la hipótesis nula (no diferencia entre las evaluaciones) es cierta, la probabilidad del valor encontrado ha de ser $\leq 5\%$, ¿cuál es el valor límite que puede llegar a tener z , a partir del cual vamos a rechazar que la diferencia entre las dos evaluaciones es cero?; por otra parte, ¿cómo sabemos el nivel de significación que corresponde al valor de z que hayamos podido encontrar? Para responder a estas preguntas hay que recurrir a la tabla de probabilidades de la distribución normal, pero antes hemos de decidir si utilizamos las probabilidades de una cola o de dos colas.

4.3. Buscar significación estadística de una cola

La distribución normal es, como es sabido, perfectamente simétrica en torno al valor central que es cero y tiene dos extremos o colas, de forma que cuando se buscan niveles de significación (probabilidades de que el valor real sea cero) para los correspondientes valores de z pueden considerarse una de las dos colas o las dos simultáneamente; de hecho hay tablas que dan los niveles de significación de una u otra forma; el mirar sólo una cola o las dos simultáneamente tiene relación con la formulación de la hipótesis alternativa, (tal como se representa en la figura 10.1), y repercute en los niveles de significación en el sentido de que mirando la significación en sólo una cola va a ser mayor para un mismo valor de z , de forma que hay incluso algunos valores de z que son significativos si los vemos en una cola, pero no lo son si se miran en las dos simultáneamente. El test de dos colas es más restrictivo.

¿Por qué podemos mirar la significación sólo en un lado o cola? Ya hemos mencionado que ello está en relación con la formulación de la hipótesis alternativa. Nuestra hipótesis alternativa es que hemos mejorado (diferencia positiva); esta hipótesis es unidireccional y sólo nos interesa la significación de valores de z positivos. Si nuestra pregunta no fuese ¿hemos mejorado? sino, ¿hay diferencia entre las dos evaluaciones? estaríamos buscando la significación de diferencias tanto positivas como negativas, y por tanto debemos considerar ambas colas simultáneamente. Este segun-

do razonamiento bidireccional está en la base de la estimación del intervalo de confianza de la diferencia, que veíamos más arriba. Sin embargo, el que nuestra hipótesis sea claramente unidireccional (en el sentido de la mejora) y ser el test de una cola menos restrictivo, es lo que nos hace sugerir como método más conveniente para ver si la diferencia entre las dos evaluaciones es significativa el test del valor de z de una cola (en el sentido de la mejora).

Vamos a ilustrar todos estos razonamientos y procedimientos aplicándolo a los criterios 1 y 3 del cuadro 10.4.

Para el criterio 1, ya vimos que el intervalo de confianza de la diferencia no contenía el valor cero, y que por tanto podíamos concluir que la diferencia no era debida al azar o los errores del muestreo; aplicando el test estadístico vamos a ver cuál es el nivel de significación de la diferencia encontrada. Para aplicar la fórmula del valor de z , sólo nos falta calcular la proporción conjunta que sería, puesto que las muestras en las dos evaluaciones son iguales,

$$\frac{P_2 + P_1}{2} = \frac{0.95 + 0.75}{2} = 0.85$$

(si las dos muestras no hubiesen sido de igual tamaño, calculamos la proporción conjunta de la siguiente manera: $\frac{n_2 \cdot P_2 + n_1 \cdot P_1}{n_1 + n_2}$).

El valor de z sería:

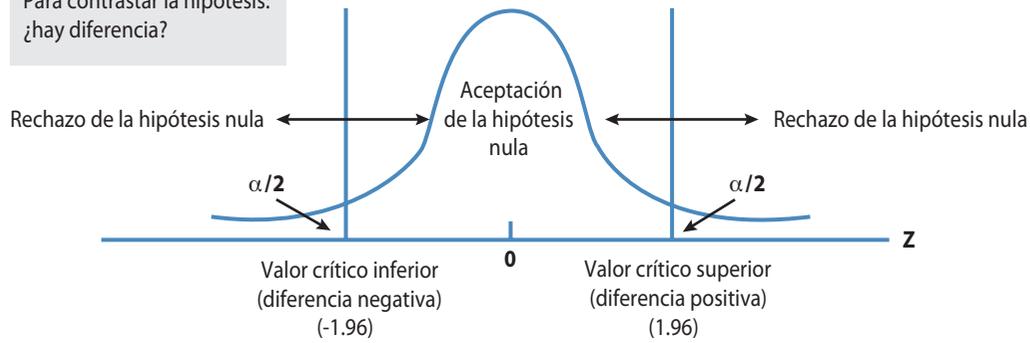
$$z = \frac{P_2 - P_1}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.95 - 0.75}{\sqrt{(0.85)(0.15) \left(\frac{1}{60} + \frac{1}{60} \right)}} = 3.07$$

Para ver el nivel de significación buscamos en la tabla de la distribución normal (una cola), que reproducimos como cuadro 10.5, la probabilidad que corresponde a ese valor de z . Para ello buscamos en la primera columna el valor de las unidades y primer decimal (3.0) y en la fila del encabezamiento el segundo decimal (0.07), el lugar donde se cruzan nos da la probabilidad de encontrar por muestreo ese valor de z si la diferencia real entre las dos evaluaciones fuese cero; este es el nivel de significación del test. El valor que encontramos en la tabla es < 0.001 . Muy significativo. Podemos afirmar que la diferencia (mejora) es real, con un riesgo (probabilidad) de equivocarnos < 0.001 .

La significación del valor de z la buscamos en una cola de la distribución normal, tal como corresponde a nuestra hipótesis alternativa (¿hemos mejorado?) que supone y espera diferencias sólo en sentido positivo.

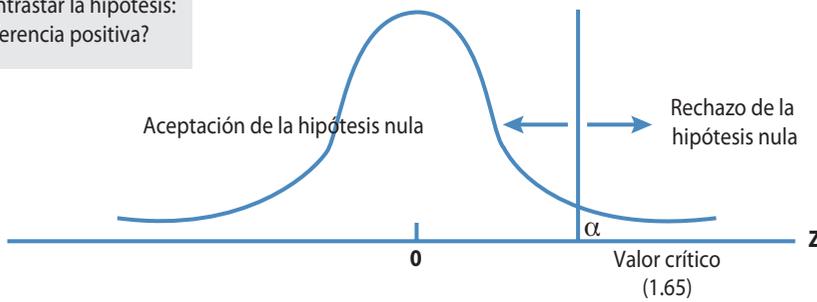
1. Test de dos colas

Para contrastar la hipótesis:
¿hay diferencia?



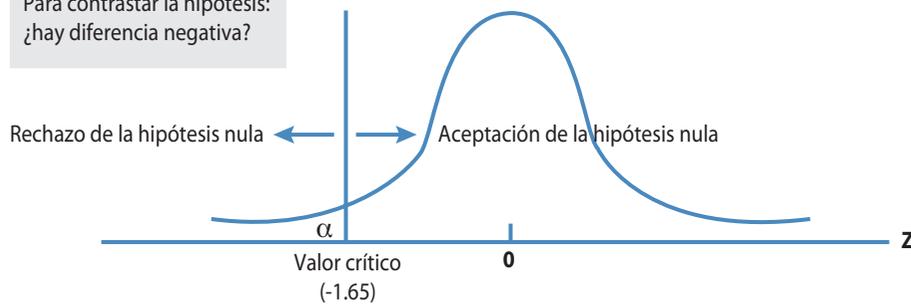
2. Test de una cola

Para contrastar la hipótesis:
¿hay diferencia positiva?



3. Test de una cola

Para contrastar la hipótesis:
¿hay diferencia negativa?



* La hipótesis nula es en todos los casos que no hay diferencia. Los valores críticos están establecidos para una significación de 0,05 (5%)
Fuente: Adaptado de Richards LD, La Cava JJ.

Figura 10.1. Test de hipótesis de la diferencia entre dos evaluaciones. Significación (probabilidades) de una o dos colas según la formulación de la hipótesis alternativa*

Para el criterio 3, $p_2=0.98$; $p_1=0.9$;

$$p = \frac{0.98 + 0.9}{2} = 0.94$$

$$z = \frac{0.98 - 0.9}{\sqrt{(0.94)(0.06) \left(\frac{1}{60} + \frac{1}{60} \right)}} = 1.85$$

buscando en la tabla el lugar donde se cruzan 1.8 y 0.05 vemos que la probabilidad de este valor de z, si no hubiese diferencia en la realidad, es 0.032; como es ≤ 0.05 (nuestro límite de riesgo para decidir), rechazamos la hipótesis nula y diremos que la diferencia entre las dos evaluaciones es estadísticamente significativa.

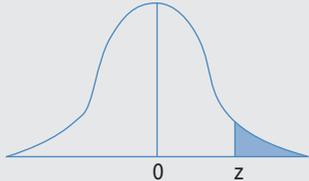
Si utilizamos para este mismo criterio el procedimiento de ver el intervalo de confianza de la diferencia tendríamos:

$$(p_2 - p_1) \pm 1.96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = 0.08 \pm 1.96 \sqrt{\frac{(0.9)(0.1)}{60} + \frac{(0.98)(0.02)}{60}} = 0.08 \pm 0.084$$

Lo que da unos límites de -0.004 y 0.164 , que contienen el valor cero, con lo que concluiríamos que la diferencia puede ser debida al azar. Esta discrepancia con el otro proceder es debida a que el intervalo de confianza es bidireccional, mientras que con el test estadístico hemos podido comprobar nuestra hipótesis de mejora que es unidireccional, de una cola; queda así de manifiesto la mayor conveniencia de utilizar el test del valor de z,

Tanto el intervalo de confianza de la diferencia como un test estadístico de dos colas son más restrictivos, y por tanto menos convenientes, para detectar diferencias significativas de mejora.

Cuadro 10.5
Probabilidades de un extremo o cola de la curva estandar normal



Esta tabla da el valor de la zona sombreada

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1.5	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058	0.057	0.056
1.6	0.055	0.054	0.053	0.052	0.051	0.049	0.048	0.048	0.046	0.046
1.7	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.038	0.037
1.8	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.031	0.030	0.029
1.9	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024	0.024	0.023
2.0	0.023	0.022	0.022	0.021	0.021	0.020	0.020	0.019	0.019	0.018
2.1	0.018	0.017	0.017	0.017	0.016	0.016	0.015	0.015	0.015	0.014
2.2	0.014	0.014	0.013	0.013	0.013	0.012	0.012	0.012	0.011	0.011
2.3	0.011	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.008
2.4	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.006
2.5	0.006	0.006	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005
2.6	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
2.7	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
2.8	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
2.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001
3.0	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

- Valores de $z > 1,65$ tienen una probabilidad (significación) $< 0,05$.

- Valores de $z > 3,0$ tienen una probabilidad (significación) $< 0,001$

Fuente: Adaptado de Colton T., Estadística en medicina, Ed.Salvat, Barcelona, 1989.

para no clasificar como no significativas diferencias que sí lo son. De igual manera, no sería conveniente utilizar un test estadístico de dos colas.

Como regla general, tal como puede verse en el cuadro 10.5, valores de $z \geq 1.65$ (valor que hemos marcado en la tabla) nos van a indicar que la mejora entre las dos evaluaciones es significativa (error $\alpha < 0.05$).

Si encontramos diferencias negativas entre la segunda y primera evaluación no merece la pena que calculemos significación estadística: como lo que vamos buscando es exclusivamente mejorar, en ese criterio está claro que no lo hemos conseguido.

El cuadro 10.6 contiene el análisis estadístico completo de los criterios del cuadro 10.2, comparando la primera y segunda evaluación. Este análisis numérico debe completarse con una presentación y análisis gráfico adecuado, sea sobre la base de los niveles de cumplimiento o analizando lo mejorado y lo que queda por mejorar con un Pareto antes-después.

5. Representación gráfica de las diferencias en los niveles de calidad entre dos evaluaciones

Al igual que veíamos en el capítulo 7, para la representación gráfica de los resultados de una evaluación, en la representación de los resultados conjuntos de dos evaluaciones podemos estar interesados en subrayar las mejoras en los niveles de cumplimiento de cada criterio, o bien la disminución del número de defectos (incumplimientos) encontrados. De igual forma, podemos querer representar las estimaciones puntuales o los intervalos de confianza. En todos los casos, los gráficos más útiles son los que ya veíamos en el capítulo 7, solo que añadiendo los resultados de la segunda evaluación en la forma que resulte más útil e informativa.

En la figura 10.2 se resumen los diversos objetivos de la comparación y representación gráfica correspondiente. Comencemos por revisar aquellas que son útiles para mostrar diferencias en el nivel de calidad, ejemplificando su caso con datos del cuadro 10.4.

La representación gráfica de las diferencias del nivel de calidad completa el análisis comparativo de las dos evaluaciones.

Cuadro 10.6

Evaluación de la calidad del diagnóstico, anamnesis y exploración física inicial del paciente hipertenso antes y después de la intervención para mejorar. Diferencias en los niveles de cumplimiento de los criterios en muestras aleatorias de 60 casos

Criterio*	1ª Evaluación p ₁ (IC 95%)		2ª Evaluación p ₂ (IC 95%)		Mejora absoluta p ₂ - p ₁	Mejora relativa $\frac{p_2 - p_1}{1 - p_2}$	Significación estadística p
1. Diagnóstico correcto	75.0	(±11.0)	95.0	(±5.5)	20	80%	<0.001
2. Antecedentes familiares	80.0	(±10.1)	86.0	(±8.8)	6	30%	NS
3. Antecedentes personales	90.0	(±7.6)	98.0	(±3.5)	8	80%	0.046
4. Consumo de tabaco	30.0	(±11.6)	70.0	(±11.6)	40	57%	<0.001
5. Consumo de alcohol	10.0	(±7.6)	50.0	(±12.7)	40	44%	<0.001
6. Peso y talla y/o IMC	85.0	(±9.0)	84.0	(±9.3)	—	<0.001	—
7. Auscultación cardiaca	75.0	(±11.0)	80.0	(±10.1)	5	20%	NS
8. Exploración abdominal	40.0	(±12.4)	57.0	(±12.5)	17	28.3%	0.031
9. Pulsos periféricos	50.0	(±12.7)	60.0	(±12.4)	10	20%	NS
10. Soplos carotídeos	55.0	(±12.6)	75.0	(±11.0)	20	44%	<0.001
11. Edemas	25.0	(±11.0)	50.0	(±12.7)	25	33%	<0.001
12. Fondo de ojo	15.0	(±9.0)	25.0	(±11.0)	10	11.8%	NS

* Este es el enunciado. La descripción detallada de los criterios debe seguir el formato indicado en el capítulo 4.

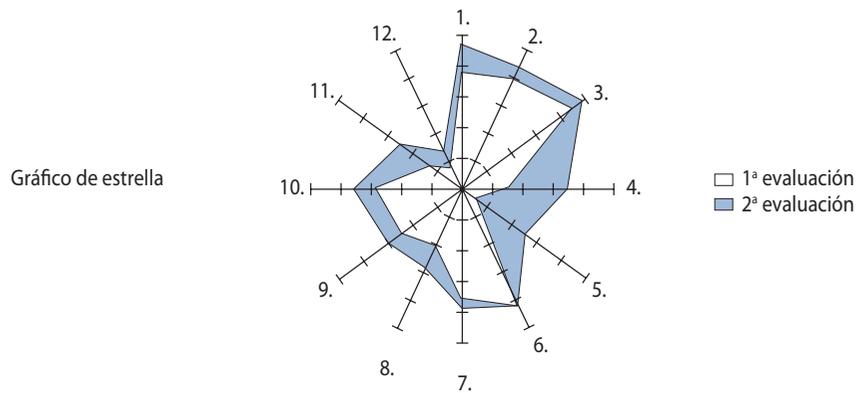
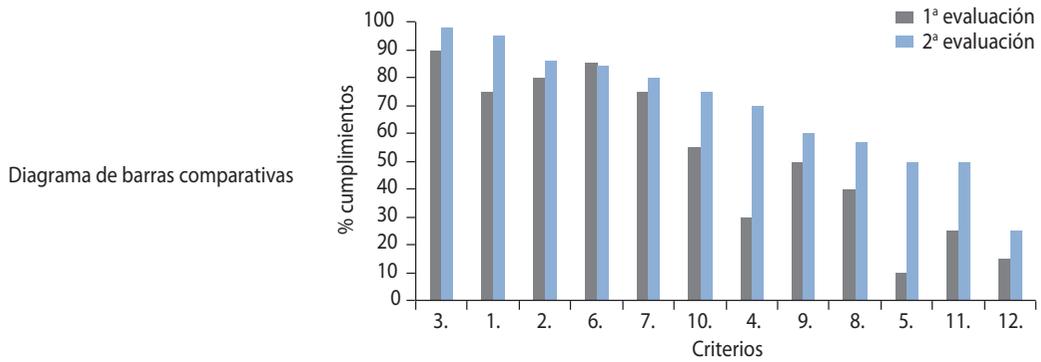
p₁ = cumplimiento en la primera evaluación.

p₂ = cumplimiento en la segunda evaluación.

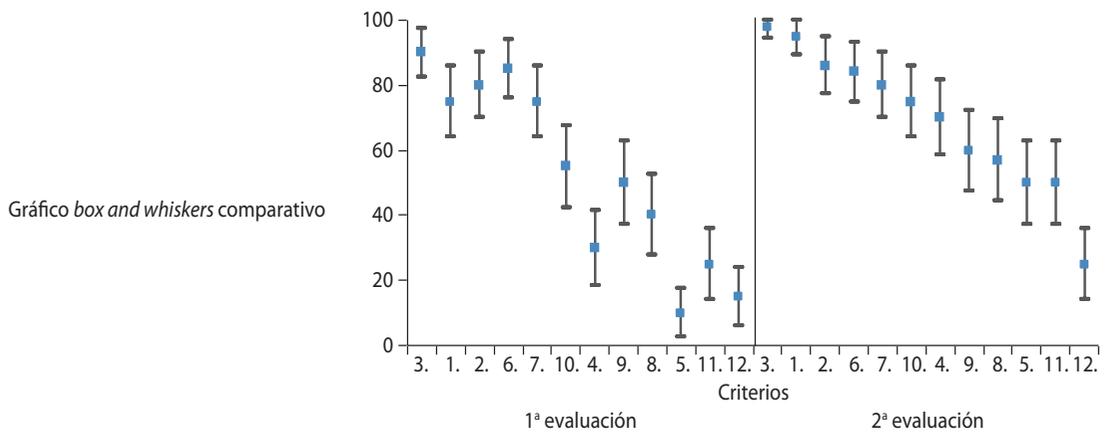
NS: No significativa ($p > 0.05$).

1. Representación de diferencias en el nivel de calidad

Diferencias en estimaciones puntuales



Diferencias con intervalos de confianza



2. Análisis de la mejora y defectos por mejorar

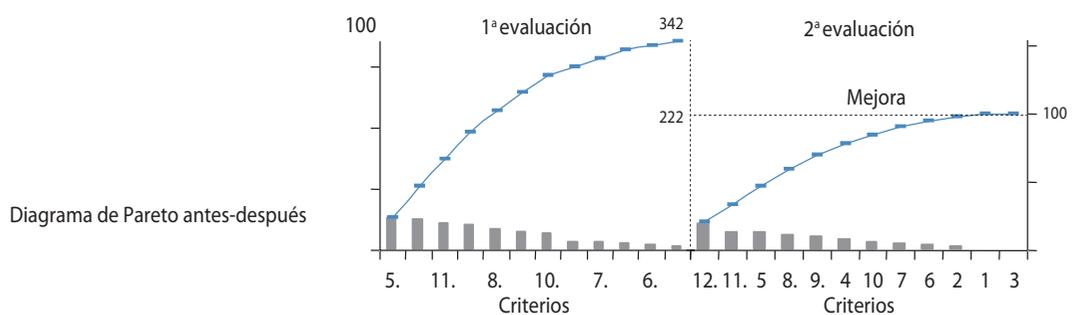


Figura 10.2. Representación y análisis gráfico de la comparación entre dos evaluaciones. Resumen de objetivos y presentación

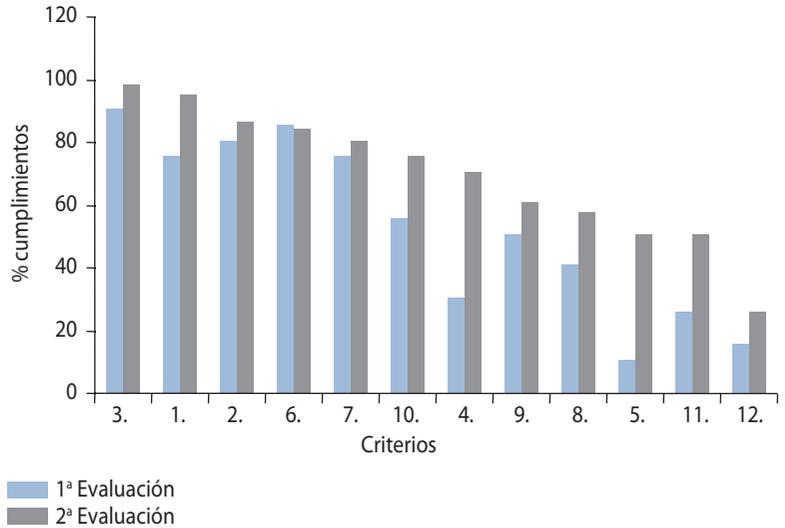
5.1. Representación gráfica de las diferencias en estimaciones puntuales del nivel de cumplimiento

Para este objetivo las dos opciones más útiles son el diagrama de barras comparativas y el gráfico de estrella. En el gráfico de barras lo que hacemos es yuxtaponer la barra correspondiente al nivel de cumplimiento en la 2ª evaluación; a lo que teníamos en la 1ª evaluación. El orden de las barras no tiene porqué ser el de la lista de criterios. De hecho es preferible o bien mantener el orden que había en la primera evaluación (de más cumplimiento, a menos cumplimiento) o establecer el orden que haya resultado en la segunda evaluación (también de más cumplimiento a menos cumplimiento). Esta segunda opción tiene la ventaja de evidenciar mejor la situación actual, y es la que hemos utilizado para representar los datos del cuadro 10.4 en el gráfico de la figura 10.3.

La representación en gráfico de estrella o radar, tiene la ventaja de poder visualizar también el área de mejora global, correspondiente al área entre las líneas que representan las dos evaluaciones. Al mismo tiempo es igualmente visible, incluso mejor que en el gráfico de barras, hasta qué punto se ha mejorado en cada criterio individualmente y lo que queda por mejorar en cada uno. Es pues un gráfico más informativo, cuya única desventaja es, como ya dijimos en el capítulo 7, que pierde en claridad, sobre todo en la comparación criterio a criterio, a medida que aumenta el número de criterios representados. La figura 10.4 es el gráfico de estrella comparando las dos evaluaciones del cuadro 10.4.

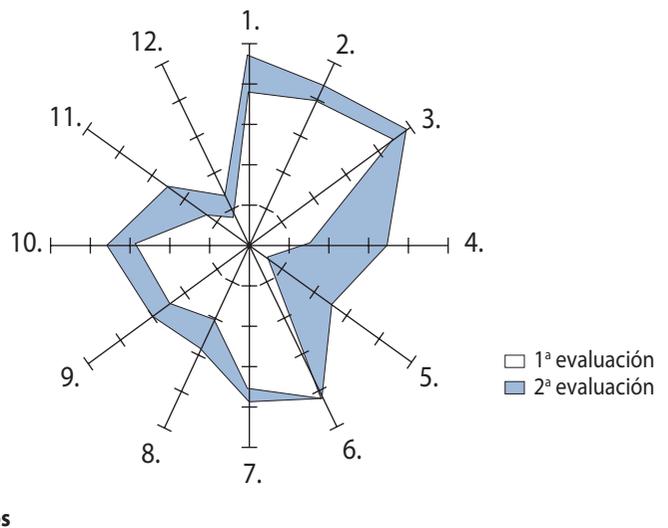
5.2. Representación gráfica de diferencias en intervalos de confianza

No es frecuente analizar y representar esta información. Lo más normal es representar las estimaciones puntuales (barras, estrella) o defectos (Pareto) e indicar, si acaso, en qué criterio es la diferencia estadísticamente significativa. No obstante, si quisiésemos plasmar gráficamente las diferencias en las estimaciones de intervalos de confianza, el gráfico de



Criterios: 1. Diagnóstico correcto. 2. Antecedentes familiares. 3. Antecedentes personales. 4. Consumo de tabaco. 5. Consumo de alcohol. 6. Peso y talla y/o IMC. 7. Auscultación cardiaca. 8. Exploración abdominal. 9. Pulsos periféricos. 10. Soplos carotídeos. 11. Edemas. 12. Fondo de ojo.

Figura 10.3. Comparación gráfica de dos evaluaciones. Gráfico de barras



Criterios

- | | |
|----------------------------|--------------------------|
| 1. Diagnóstico correcto | 7. Auscultación cardiaca |
| 2. Antecedentes familiares | 8. Exploración abdominal |
| 3. Antecedentes personales | 9. Pulsos periféricos |
| 4. Consumo de tabaco | 10. Soplos carotídeos |
| 5. Consumo de alcohol | 11. Edemas |
| 6. Peso y talla y/o IMC | 12. Fondo de ojo |

Figura 10.4. Comparación gráfica de dos evaluaciones. Gráfico de estrella

elección sería el de tipo *box and whiskers*, cuya explicación vimos también en el capítulo 7.

Como en el gráfico de barras, puede ser útil ordenar los criterios de más a menor nivel de cumplimiento según los resultados de la segunda evaluación.

Cuando hay muchos criterios, puede también ser visualmente más claro si prescindimos de la *caja* y representamos sólo la estimación puntual y los límites del intervalo de confianza del 95%. En la figura 10.5 se han representado los datos del cuadro 10.4 con un gráfico *box and whiskers* en el que hemos introducido las dos modificaciones mencionadas (ordenar los criterios según los resultados de la segunda evaluación y representar sólo la estimación puntual y los límites del intervalo de confianza).

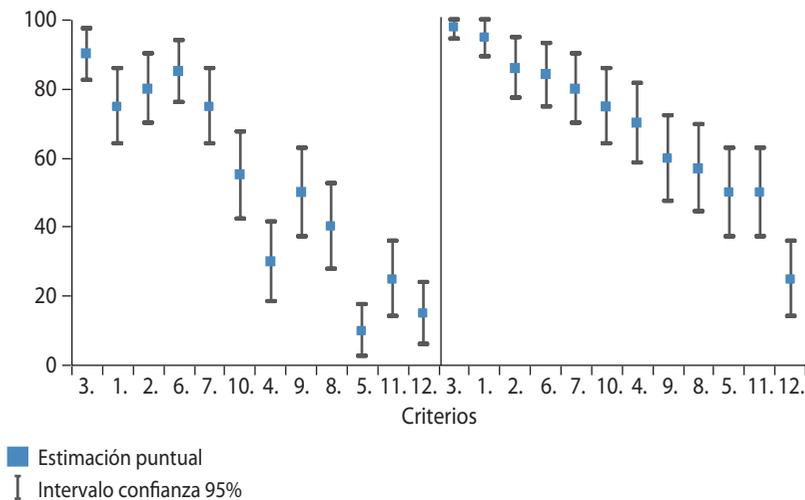
Con todo, reviste un mayor interés analizar explícitamente no sólo la mejora conseguida sino también los defectos por mejorar. La mejor forma de hacerlo es con un diagrama de Pareto antes-después, una forma de representar la diferencia entre las dos evaluaciones que ofrece más información que ningún otro gráfi-

co, y que se convierte así en la representación gráfica de elección, caso de querer seleccionar sólo una.

6. Análisis gráfico de la mejora y defectos por mejorar: diagrama de Pareto antes-después

En el capítulo 7 se revisó paso a paso la construcción de un diagrama de Pareto de los defectos como el tipo de análisis gráfico de más utilidad para priorizar hacia dónde dirigir las acciones para la mejora. A la hora de comparar la reevaluación con la situación de partida también nos interesa, aparte de ver lo que hemos mejorado, tener una imagen clara de las características de lo que queda por mejorar. Para conjuntar ambas imágenes (mejora conseguida y mejora por conseguir) nada mejor que un diagrama de Pareto antes-después elaborado de una manera determinada.

En los manuales sobre métodos para el control y la mejora de la calidad se pueden encontrar varias formas de diagrama de Pa-



Criterios

- | | |
|----------------------------|--------------------------|
| 1. Diagnóstico correcto | 7. Auscultación cardiaca |
| 2. Antecedentes familiares | 8. Exploración abdominal |
| 3. Antecedentes personales | 9. Pulsos periféricos |
| 4. Consumo de tabaco | 10. Soplos carotídeos |
| 5. Consumo de alcohol | 11. Edemas |
| 6. Peso y talla y/o IMC | 12. Fondo de ojo |

Figura 10.5. Comparación gráfica de dos evaluaciones *box and whiskers* antes-después

reto antes-después, siendo probablemente lo más frecuente la simple yuxtaposición de los Paretos de ambas evaluaciones construidos de forma independiente, cada uno con los datos sobre los defectos de su correspondiente evaluación; en otras ocasiones aparecen las barras apiladas, en vez de en secuencia horizontal, y yuxtapuestas por su base tomando la forma de lo que se llama “pirámide de Pareto”. Sin embargo ninguna de estas formas es visualmente tan clara e informativa como la que se consigue introduciendo una pequeña modificación: *mantener en el Pareto de la segunda evaluación el tamaño y escala que tenía en la primera evaluación el eje que representa el número absoluto de defectos*; si hemos logrado mejorar, el Pareto de la segunda evaluación se representará en una parte de este eje, de forma que la diferencia aparece claramente como mejora conseguida. Al mismo tiempo, el diagrama de la 2ª evaluación puede interpretarse aisladamente, como cualquier Pareto, para identificar qué hacer para seguir mejorando; adicionalmente, al mantenerse la misma escala, la diferencia en cada criterio puede compararse individualmente viendo las barras correspondientes de la primera y segunda evaluación.

Así pues, construido como proponemos, el Pareto antes-después permite:

1. *Evidenciar la magnitud* de la mejora global conseguida, expresada por la diferencia entre el número total de defectos (incumplimientos) de la primera y segunda evaluación, y el área que delimita en el Pareto de la 2ª evaluación.
2. *Analizar y priorizar* lo que nos queda que mejorar, y qué criterios son los principales causantes de los defectos de calidad.
3. *Comparar la mejora* (disminución del número de incumplimientos) de cada criterio individualmente.

La forma más conveniente de este tipo de Pareto antes-después es probablemente la que se representa en el esquema de la figura 10.2. Esta es una de las dos formas de representación que han surgido en nuestra experiencia, modificando la forma más común que sería la de construir dos Paretos independientes, dado que ambos Paretos tienen un eje común que ha de ser idéntico (el correspondiente al número

absoluto de defectos o incumplimientos). En estas dos formas alternativas, representadas en la figura 10.6 junto a la más común, se unen los dos diagramas en uno solo con tres ejes, de los cuales el central es el del número absoluto de defectos (incumplimientos), obligatoriamente igual para los dos diagramas, mientras que el de la izquierda es el de la frecuencia relativa (% de incumplimientos) para la primera evaluación y el de la derecha es el de la frecuencia relativa en la segunda evaluación. Con este esquema de base, tanto las barras del diagrama como la curva de porcentaje acumulado para identificar los pocos vitales pueden partir del eje central en ambas evaluaciones (con lo cual el diagrama de la primera evaluación resultará ordenado de derecha a izquierda, figura 10.6c), o mantener el orden de izquierda a derecha para los dos (figura 10.6b). En todos los casos van a ser representaciones gráficas informativas de la diferencia entre las dos evaluaciones y de las características de lo que aún queda por mejorar. Depende de las preferencias de cada uno elegir una u otra de las tres modalidades.

Los datos del cuadro 10.4, convertidos en información útil para los diagramas de Pareto son los que figuran en el cuadro 10.7. Con ellos hemos construido los tres diagramas antes-después de la figura 10.6, donde se aprecia a simple vista el área de mejora, a la vez que el Pareto de después es interpretable de forma independiente como ayuda para decidir qué hacemos a continuación.

En caso de que la muestra empleada en la segunda evaluación no sea igual que la de la primera hay que hacer un pequeño ajuste para poder utilizar cualquiera de las tres modalidades de Pareto antes-después que hemos propuesto. El ajuste consiste en calcular, por simple regla de tres, lo que supondría el número de defectos encontrados en la segunda evaluación, si el tamaño de muestra fuese idéntico al de la primera evaluación. Por ejemplo, si los datos de del cuadro 10.5 correspondieran a una primera evaluación de 60 casos y una segunda de 80, no podríamos utilizarlos directamente para representar el diagrama de Pareto antes-después que proponemos, dado que en 80 casos vamos a encontrar un mayor número de defectos (incumplimientos) para proporciones de cumplimiento idénticos en las dos evalua-

La representación gráfica más completa e informativa es el gráfico de Pareto antes-después.

Para que el Pareto antes-después resulte más informativo hay que mantener constante en los dos diagramas el eje del número absoluto de defectos.

Si la muestra no es idéntica en las dos evaluaciones, hay que ajustar el número de defectos (incumplimientos) de la segunda evaluación tomando como referencia el tamaño de muestra de la primera evaluación.

Cuadro 10.7

Comparación de dos evaluaciones, frecuencia absoluta, relativa y acumulada de los incumplimientos de los criterios evaluados

Criterio	1ª evaluación			Criterio	2ª evaluación		
	Nº incum (frecuencia absoluta)	% (Frecuencia relativa)	Frecuencia acumulada		Nº incum (frecuencia absoluta)	% (Frecuencia relativa)	Frecuencia acumulada
5. Consumo de alcohol	54	15.8	15.8	12. Fondo de ojo	45	20.3	20.3
12. Fondo de ojo	51	14.9	30.7	11. Edemas	30	13.5	33.8
11. Edemas	45	13.2	43.9	5. Consumo alcohol	30	13.5	47.3
4. Consumo tabaco	42	12.3	56.2	8. Exploración abdominal	26	11.7	59.0
8. Exploración abdominal	36	10.5	66.7	9. Pulsos periféricos	24	10.8	69.8
9. Pulsos periféricos	30	8.8	75.4	4. Consumo tabaco	18	8.1	78.0
10. Soplos carotídeos	27	7.9	83.3	10. Soplos carotídeos	15	6.8	84.7
1. Diagnostico correcto	15	4.4	87.7	7. Auscultación cardiaca	12	5.4	90.1
7. Auscultación cardiaca	15	4.4	92.1	6. Peso y talla y/o IMC	10	4.5	94.6
2. Antecedentes familiares	12	3.5	95.6	2. Antecedentes familiares	8	3.6	98.2
6. Peso y talla y/o IMC	9	2.6	98.3	1. Diagnostico correcto	3	1.4	99.6
3. Antecedentes personales	6	1.8	100.0	3. Antecedentes personales	1	0.5	100.0
	342	100			222	100	

ciones; en consecuencia no podríamos utilizar el "truco" principal de nuestro diagrama conjunto que consiste, como hemos visto, en mantener común el eje de número absoluto de defectos y realizar la comparación en referencia a él. Para solucionarlo, normalizamos los datos de la segunda evaluación, tomando como referencia la primera evaluación, y calculamos el número de defectos que corresponderían a los encontrados en la 2ª evaluación si el tamaño de muestra fuese como el de la primera evaluación. De la siguiente forma:

$$\frac{\text{defectos en la 2ª evaluación } (d_2)}{\text{muestra en la 2ª evaluación } (n_2)} = \frac{x}{\text{muestra en la 1ª evaluación } (n_1)}$$

$$\text{de donde } x = \frac{d_2 \cdot n_1}{n_2},$$

cálculo que realizaríamos para cada criterio, de forma que el número absoluto de defectos sea perfectamente comparable a los que había en la primera evaluación.

En nuestros datos, por ejemplo, el número de defectos que había que utilizar para el cri-

terio 12 en Pareto comparativo de la 2ª evaluación, si n_2 fuese 80 (y no 60 como n_1) sería:

$$x = \frac{45 \cdot 60}{80} = 34$$

de forma que los 45 incumplimientos de la muestra de 80 casos contarían como 34 para la comparación con la 1ª evaluación que se hizo sobre 60 casos.

Con los números absolutos de defectos ajustados de esta forma para todos los criterios, construiremos nuestro Pareto comparativo.

7. ¿Qué hacemos a continuación? Cursos de acción tras la reevaluación

En general, dentro de las actividades para mejorar la calidad de la atención que ofrecemos a nuestra población, siempre tendremos la opción de iniciar nuevos ciclos de mejora, diseñar o planificar mejor los servicios que ofrecemos, y construir y monitorizar indicadores relevantes y representativos de los aspectos que más nos interesen de nuestro trabajo. Estos son los

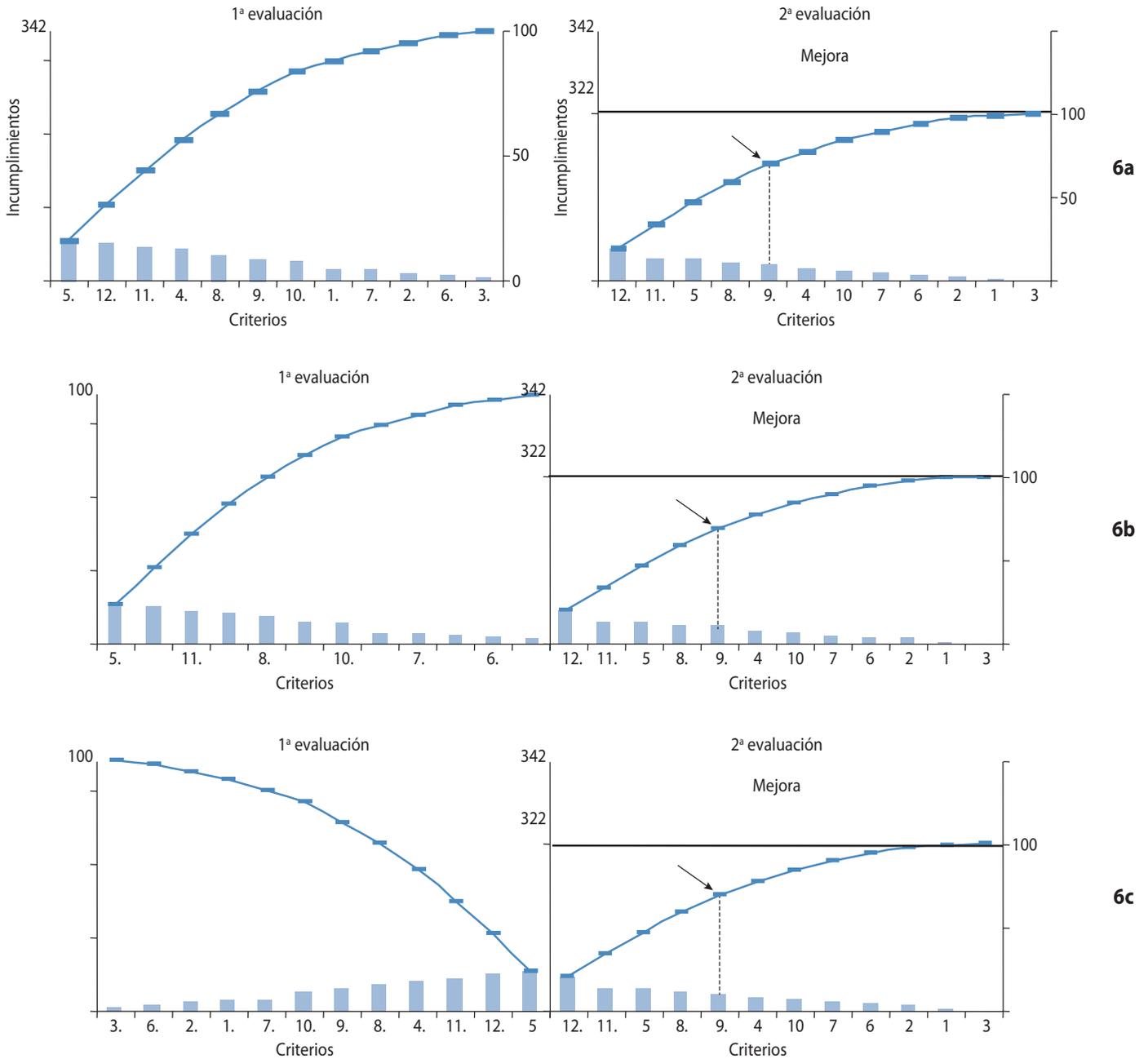


Figura 10.6. Comparación gráfica de dos evaluaciones

tres grupos de actividades que mencionábamos en el capítulo 1, uno de los cuales (Ciclos de mejora) se describe paso a paso en este manual. Los otros dos (monitorización y diseño o planificación) son objeto de otros manuales que acompañan a este.

Sin embargo, en cualquier caso hay que decidir qué hacemos en relación con el ciclo de mejora tras analizar la reevaluación con la metodología descrita en este capítulo. Obviamente, lo que decidamos hacer va a estar condicionado por el resultado obtenido en el ciclo

Si no hemos mejorado, hay que reiniciar el ciclo y pensar en qué podemos haber fallado.

de mejora. En primer lugar, si no hemos conseguido mejorar hasta los niveles que queremos, y debemos, estar, no podemos dar por concluido el ciclo de mejora. Es muy probable que nos hayamos equivocado al analizar las causas o al diseñar la intervención, que, obviamente, no ha sido efectiva si no hemos logrado mejorar. Lo que debemos hacer entonces es reiniciar el ciclo, o al menos retomarlo en aquellos pasos que pensamos son el origen de que no hayamos mejorado (¿análisis de causas?; ¿diseño de criterios?, ¿diseño de intervención?).

Caso de haber logrado mejorar satisfactoriamente, tenemos dos opciones ante nosotros. Una es dejar el tema como está y otra diseñar un plan de monitorización para asegurarnos que la mejora conseguida se mantiene. La primera opción, dejar el tema como está, sería lógica si los niveles de calidad alcanzados son

óptimos y la intervención ha sido lo suficientemente robusta, generalmente de tipo estructural, como para pensar que va a ser difícil que volvamos a los niveles de calidad deficiente que teníamos al comenzar el Ciclo de mejora. La segunda opción supone elegir como indicador o indicadores algunos de los criterios empleados en la evaluación (o construir indicadores relacionados) y volverlos a medir de forma planificada con los métodos y planes de monitorización óptimos que nos alerten de la presencia, si se produce, de deterioro en el nivel de calidad logrado. En esta opción, para ser prácticos y operativos nos conviene conocer y practicar métodos de monitorización lo más eficientes posible, lo cual constituye otro bloque de conocimientos que debemos adquirir, y que es objeto del segundo manual de esta serie.

BIBLIOGRAFÍA

- Gitlow H, Gitlow S, Oppenheim A, Oppenheim R. Diagnosing a process. In: Tools and methods for the improvement of quality. Boston: Irwin, 1989. Cap. 12.
- Lemeshow S, Hosmer DW, Klar J, Lwanga SK. The two-sample problem. In: Adequacy of sample size in health studies. New York: WHO/John Wiley and sons, 1992. Cap. 2.
- Richard LE, LaCava JJ. Test of hypothesis. In: Business statistics. New York: McGraw Hill, 1983. Cap. 8.

Anexo 10.1**Indicaciones para la aplicación práctica de los contenidos de este capítulo**

Los contenidos metodológicos de este capítulo son útiles para analizar y documentar estadística y gráficamente la mejora conseguida tras una intervención, en referencia a la situación basal. Se trata de que quede evidente la respuesta a las preguntas siguientes: ¿cuál es el nivel de calidad que hemos logrado tras intervenir?; ¿cuál ha sido la mejora conseguida?; ¿hemos mejorado significativamente en relación a la situación de partida?; ¿cuál ha sido la mejora relativa en cada uno de los criterios o indicadores medidos?; ¿qué debemos hacer a continuación en relación al problema de calidad evaluado? Para realizar esta práctica son necesarios los resultados obtenidos en una evaluación basal y la reevaluación, así como implementar una intervención para mejorar, del nivel de calidad del tema seleccionado por cada participante o equipo de mejora. El resumen e indicaciones y formatos a utilizar para los pasos a seguir se describen a continuación.

ANÁLISIS ESTADÍSTICO

- Al igual que ocurría con la evaluación (capítulo 7), la reevaluación de una muestra de casos proporciona una estimación puntual que debe completarse con el cálculo del intervalo de confianza.
- No obstante, el principal objetivo de la reevaluación es comprobar si se ha experimentado una mejora en el nivel de calidad, y tomar, en consecuencia, decisiones en relación con el curso de acción siguiente.
- Sin embargo, la mejora observada puede ser debida al azar cuando se han evaluado muestras (mejora aparente). En consecuencia, a menos que se haya evaluado el universo de casos, es necesario determinar si las diferencias en el cumplimiento de criterios entre ambas evaluaciones son significativas desde el punto de vista estadístico o no.
- En resumen, el análisis estadístico se desarrolla a lo largo de las siguientes etapas: 1º. Estimación de la proporción de cumplimiento puntual y del intervalo de confianza en los criterios (nivel de calidad) o indicadores, tanto en la reevaluación como en la evaluación basal; 2º. Estimación de la mejora conseguida (absoluta y relativa); y, en el caso de que se hayan evaluado muestras, 3º. Grado de significación estadística de la mejora observada.

Paso 1º. Para estimar el nivel de calidad (puntual e intervalo de confianza) se procede de forma análoga a lo descrito en el capítulo 7.

Paso 2º La estimación puntual de la mejora o el cambio experimentado en el cumplimiento de un criterio entre la reevaluación (p_2) y la evaluación (p_1) se proporcionará en términos de mejora absoluta y relativa, para cada criterio. La mejora absoluta (MA) es, simplemente, la diferencia, entre la reevaluación y la evaluación, en el nivel cumplimiento del criterio en cuestión, como se resume en la siguiente fórmula:

$$MA = p_2 - p_1$$

Esta mejora se debe reflejar también, ya que aporta más información, en relación al punto de partida o margen de mejora potencial de la 1ª evaluación, mediante el cálculo de la mejora relativa (MR) de la siguiente forma:

$$MR = \frac{p_2 - p_1}{1 - p_1}$$

Para darlo en forma de porcentaje se multiplica por 100.

Continúa.../

/... continuación

Indicaciones para la aplicación práctica de los contenidos de este capítulo

Paso 3º. Por último, cuando se trabaja con muestras, es necesario descartar la posibilidad de que la mejora observada sea, en realidad, aparente; como consecuencia del azar o error del muestreo. Ello implica determinar, mediante un contraste de hipótesis de una cola, la probabilidad de la hipótesis nula:

Hipótesis nula: $p_2 \leq p_1$ "No se ha experimentado una mejora en el nivel de calidad"

- ▶ Dicha hipótesis se puede rechazar cuando tiene una probabilidad (p) suficientemente baja (normalmente menor o igual a 0.05), para aceptar, con un error ($=p$) la hipótesis alternativa:
- ▶ Hipótesis alternativa: $p_2 > p_1$ "Se ha experimentado una mejora" (con un riesgo de error= p)
- ▶ La probabilidad de la hipótesis nula (p o error) se determinará, en el caso más frecuente de que se hayan aplicado criterios dicotómicos (cumple o no cumple), mediante el test de la z de comparación de proporciones para consultar posteriormente en la tabla de la distribución del valor z de una cola (cuadro 10.5) su valor p (probabilidad de que la hipótesis nula sea cierta).

$$z = \frac{P_2 - P_1}{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$p = \frac{c_1 + c_2}{n_1 + n_2}$$

P_2 : proporción de cumplimientos en la 2ª evaluación

P_1 : proporción e cumplimientos en la 1ª evaluación.

n_2 : tamaño de la muestra en la 2ª evaluación

n_1 : tamaño de la muestra en la 1ª evaluación.

c_1 : nº de cumplimientos del criterio en 1ª evaluación.

c_2 : nº de cumplimientos del criterio en la 2ª evaluación.

Los resultados pueden resumirse en una tabla con el formato que se adjunta a continuación:

Continúa.../

/... continuación

Indicaciones para la aplicación práctica de los contenidos de este capítulo

Formato para la presentación de los datos de la reevaluación					
Criterio	1 ^a Evaluación p1 (IC 95%)	2 ^a Evaluación p2 (IC 95%)	Mejora absoluta p2 - p1	Mejora relativa p2-p1 1-p1	Significación estadística p
1.	(±)	(±)			
2.	(±)	(±)			
3.	(±)	(±)			
4.	(±)	(±)			
5.	(±)	(±)			
6.	(±)	(±)			
7.	(±)	(±)			
8.	(±)	(±)			
9.	(±)	(±)			
10.	(±)	(±)			
.	(±)	(±)			
.	(±)	(±)			

p1= cumplimiento en la primera evaluación.

p2= cumplimiento en la segunda evaluación.

NS: No significativa (p >0.05)

/... continuación

Indicaciones para la aplicación práctica de los contenidos de este capítulo

ANÁLISIS GRÁFICO: DIAGRAMA DE PARETO ANTES-DESPUÉS

(Caso de haber medido más de un criterio o indicador)

1. Elaboración de la tabla de frecuencias absolutas y relativas de incumplimientos.

Se realizará una tabla de frecuencias de incumplimientos relativos y absolutos, en la reevaluación de cada criterio, siguiendo los pasos 1º, 2º y 3º del apartado correspondiente al "Análisis de los defectos de la calidad" que se describe en la aplicación práctica del capítulo 7 con los resultados de la 1ª evaluación.

Sin embargo, si las muestras, de la reevaluación (n_2) y de la evaluación (n_1) difieren de tamaño será necesario, previamente, ajustar las frecuencias absolutas de incumplimientos de la reevaluación de cada criterio (d_2) en base al tamaño de la muestra de la 1ª evaluación, aplicando una sencilla regla de tres:

$$x = \frac{d_2 - n_1}{n_1}$$

Esta tabla junto con la correspondiente a la 1ª evaluación son las que se emplean para elaborar el gráfico de Pareto antes-después.

2. Elaboración del gráfico

Paso 1º. Dibujar un plano cartesiano con un eje horizontal y tres subejes verticales (derecho, central e izquierdo).

Paso 2º. Representar, con los datos de la primera evaluación, un diagrama de Pareto, tal como se desarrolla en los pasos 4º, 5º y 6º del apartado correspondiente al análisis de los defectos de la calidad descritos en el capítulo 7; pero, a diferencia del anterior, el porcentaje relativo de incumplimientos corresponde al eje vertical izquierdo, mientras que la frecuencia absoluta de incumplimientos corresponde al eje vertical central. Los criterios se disponen en el eje de abscisas, en la parte situada a la izquierda del eje vertical central y en orden decreciente a su frecuencia relativa de incumplimientos.

Paso 3º. Representar a partir de los datos de la tabla de incumplimientos de la segunda evaluación, otro diagrama de Pareto (barras de altura decreciente, curva acumulada y anotación del punto correspondiente al 60-80% del total de incumplimientos). La frecuencia absoluta y relativa de incumplimientos se corresponde con los ejes verticales central y derecho respectivamente. Los criterios se representan en el eje de abscisas, en la zona situada a la derecha del eje vertical central y en orden decreciente a su frecuencia relativa de incumplimientos.

Paso 4º. Unir con una recta los puntos correspondientes al total de incumplimientos absolutos y relativos (100%) de la 2ª evaluación; ambos situados en los ejes verticales central y derecho respectivamente; y trazar una paralela, delimitada por los ejes anteriores, que pase por el punto correspondiente al total de incumplimientos absolutos de la primera evaluación.

Paso 5º. Subrayar entre los ejes verticales central y derecho la zona comprendida entre las dos rectas anteriores. Esta área corresponde al espacio de mejora experimentada.

Continúa.../

Métodos y herramientas
para la realización de ciclos de mejora
de la calidad en servicios de salud

Se terminó de imprimir en diciembre de 2015.
La edición consta de 1 000 ejemplares
y estuvo al cuidado de la
Subdirección de Comunicación Científica y Publicaciones
del Instituto Nacional de Salud Pública

