UNIDAD TEMÁTICA 12

12

DISEÑO DE INTERVENCIONES PARA MEJORAR

EMCA

Gestión de la Calidad Asistencial

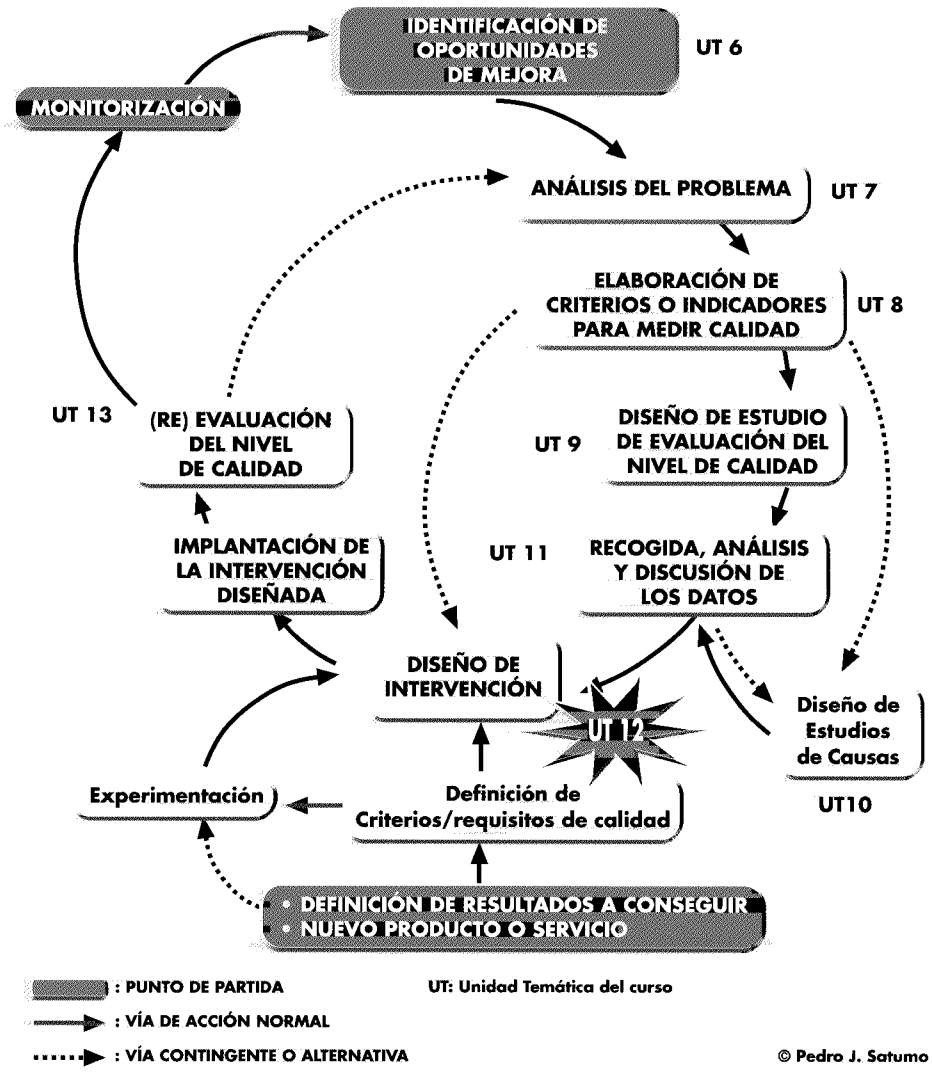
CONTENIDO GENERAL

Saber qué hacer para mejorar los aspectos evaluados y diseñar su puesta en práctica es el objetivo central de los Ciclos de Mejora. Esta UT 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, Votación múltiple) para realizar diseños participativos de una forma estructurada, y para planificar su puesta en práctica (diagrama de Gantt, diagrama de Problemas Anticipados, Carteles Narrativos).

ÍNDICE DE CONTENIDOS

OBJETIVOS ESPECÍFICOS

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.
8. Estructurar los componentes y responsabilidades de la intervención propuesta, utilizando un diagrama de Gantt.
9. Desarrollar un Cartel Narrativo del Ciclo de Mejora.
10. Analizar, estructurada y explícitamente, los problemas que puede tener la puesta en práctica de la intervención diseñada



"¡Datos, datos, datos!- gritó impacientemente-.
¡No puedo hacer ladrillos sin arcilla!".

(Sherlock Holmes)

1. INTRODUCCIÓN

Después de analizar el problema de calidad (UT 7), decidir con qué criterios vamos a medir el nivel a que nos encontramos (UT 8) y, sobre todo, tener datos que analizar y discutir (UT 11), 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, también podemos utilizar métodos estructurados que aseguren la participación de los implicados al diseñar la intervención, así como formas de realizar un seguimiento explícito de su puesta en marcha. En esta UT se explica una selección de estos métodos, efectivos pero de aplicación sencilla, para realizar un diseño participativo de la intervención y el seguimiento de una implementación. De entre los métodos útiles para un diseño participativo vamos

a destacar el *diagrama de afinidades* por su sencillez y versatilidad de aplicación en éste, sobre todo, pero también en otros pasos del ciclo de mejora; por las mismas razones, de entre los métodos de seguimiento del proceso de mejora, vamos a destacar los carteles narrativos (*story boards*), que ya vimos en la UT 6, aunque van a ser revisados también otros métodos, como material de referencia para cuando se considere oportuna su utilización.

2. DISEÑO DE INTERVENCIONES PARA MEJORAR: DIRECTRICES GENERALES

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 esta UT se revisa una selección de estos métodos

La principal virtud que pueden tener las acciones que tomemos para mejorar es precisamente 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 la UT 8 (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; (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. MEDICIÓN DE LAS CAUSAS Y CALIDAD DE LOS REGISTROS

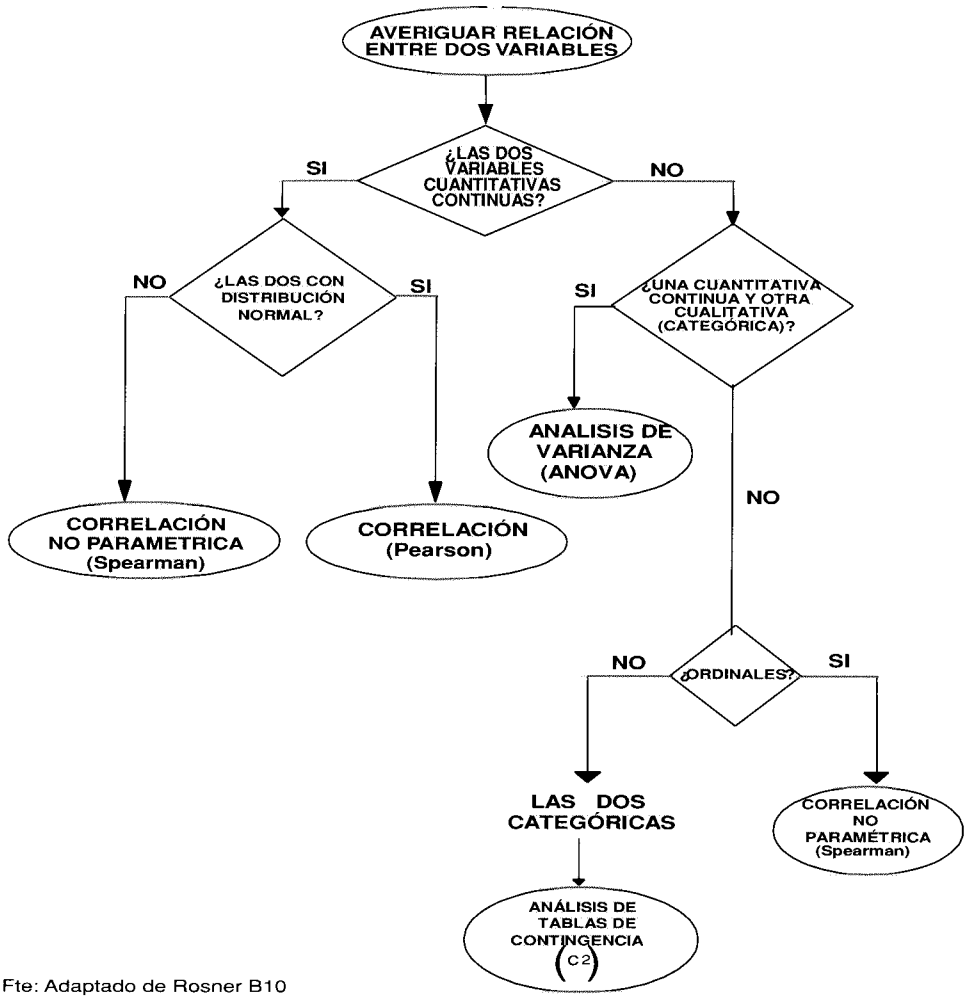
Las intervenciones son más efectivas si se basan en datos y se diseñan con la participación de los implicados en ponerlos en práctica

Hemos visto en UT previas 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, cuales 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. Siempre va a ser necesario medir al menos un criterio o indicador para saber el punto de partida y si hemos logrado o no mejorar tras la intervención. Pero 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 anteriores UT) y el *diagrama de afinidades* (que veremos más adelante), se utilizan a veces sin que vengan precedidos o acompañados de medición con datos reales de las suposiciones que los sustentan. Esto es más frecuente, como hemos dicho, 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.

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

En este tipo de estudios de causas hipotéticas, cuando hemos realizado una medición de las mismas, situación en general más conveniente que no hacerlo, lo que hay que efectuar es un estudio estadístico de su asociación o no al problema de calidad considerado. Las situaciones y test estadísticos más habituales son las que figuran en el esquema de la Figura 12.1, tal como vimos en la UT 10.

FIGURA 12.1. Test estadísticos para la asociación entre dos variables (indicador de calidad y un problema de causa)



Fte: Adaptado de Rosner B10

Sin embargo, si el trabajo práctico que hemos realizado en este curso 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 hecho en la UT 11; sobre esta base, reflexionaremos sobre la forma de que podamos hacer que se cumplan. En ocasiones, no habitualmente, podemos ver 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 que aparece al principio de cada Unidad Temática.

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: "yo lo hago correctamente pero no tengo tiempo de registrarlo", es el argumento más frecuente. ¿Cómo reaccionar ante ello?, ¿puede ser cierto que no sea la calidad de la atención lo que hayamos evaluado? Ciertamente, 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. Ese 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 la Tabla 12.1):

- **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 seguimiento, 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.
- **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 otros profesionales) preguntas y exploraciones que sólo hace falta hacer una vez. Este sería el caso de los antecedentes familiares, personales y toda las serie de exploraciones básicas que hay que realizar en caso de patologías crónicas.

TABLA 12.1. Evaluación en base a datos registrados

MÁS VÁLIDA SI:
• El registro del dato es importante para el proceso que se evalúa.
• El registro del dato evita repeticiones.

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.

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 los que se registra o no, sin asegurarse previamente la importancia de lo que se va a buscar para un determinado proceso asistencial.

2.2. DISEÑO DE INTERVENCIÓN DE FORMA PARTICIPATIVA

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.

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.

3. MÉTODOS PARA UN DISEÑO PARTICIPATIVO

La Tabla 12.2 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 otras UT (el *análisis de campos de fuerzas* en la UT 5, y la *técnica de Grupo Nominal* en la UT 6). 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 en este sentido el objetivo del análisis para el diagrama de campos de fuerzas, o hacer de la propuesta de intervenciones el objetivo de reflexión para el Grupo Nominal. Los otros métodos mencionados en la Tabla 12.2, 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 medición empírica, tal como vimos en la UT 10. Probablemente el *diagrama de afinidades* es, junto al *análisis de campos de fuerza*, uno de los métodos más genuinamente útiles para diseñar estrategias e intervenciones, y es conveniente destacarlo sobre los demás. Vamos a recordar brevemente en que consiste la encuesta sobre causas y la votación múltiple, y revisar más en detalle el diagrama de afinidades (el análisis de campos de fuerza y la técnica del Grupo Nominal, fueron vistos en detalle en las UT 5 y 6), aunque recordemos de nuevo que no va a resultar siempre *obligatorio* utilizarlos; es muy frecuente, en el caso de los criterios clínicos, que la intervención para mejorar resulte evidente sólo a la vista y discusión de los datos de la evaluación.

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.

TABLA 12.2. 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.

3.1. ENCUESTA SOBRE ACTIVIDADES 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 causas y 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.

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 siguientes:

1. Preparar el listado de acciones.
2. Tras examinarlo, cada persona implicada entrega una lista en la que ha seleccionado un tercio de la lista (por ejemplo, si el listado tiene 30, selecciona 10; si tiene 15, selecciona 5, etc.).
3. Se cuentan los "votos", viendo el número de menciones para cada acción de la lista, y se eliminan aquellas con menos menciones, reduciendo la lista. El número mínimo de "votos" para mantener la acción 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 es lo suficiente reducido e importante como para constituir la base del diseño de intervención.

Los métodos de consenso que no manejan datos tienen el riesgo de equivocarse si las percepciones del grupo no son las correctas

La votación múltiple sin que se base en mediciones previas 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 en base a 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. 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 a 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 (resumidos en la Tabla 12.3):

TABLA 12.3. Construcción del 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

- **Planteamiento del Problema.** 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 del diagnostico y exploración inicial del paciente hipertenso ¿qué podemos hacer para mejorar?

- **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 de tamaño 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:

  a) 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 cuál cada idea debe exponerse con un mínimo de tres palabras, que incluyan preferentemente un sujeto y 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 que el criterio sobre exploración de fondo de ojo 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 además 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 "si" 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 creer que cuando estén en blanco quiere decir que no fuma o no bebe, además de que nos da pié para completar la información sobre hábitos tóxicos (tabaco y alcohol) si señalamos la casilla del "si".

El diagrama de afinidades es una técnica grupal útil para recoger, ordenar y convertir en plan de acción todas las ideas del grupo para mejorar el problema evaluado. Para ello hay que seguir unos pasos perfectamente estructurados.

b) Utilizar papel autoadhesivo, fácil de pegar y despegar para realizar posteriormente las agrupaciones y desagrupaciones 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.

- **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.
- **Agrupación de las acciones en grupos lógicos**, 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 hay ideas duplicadas o muy semejantes, se funden en una sola.
  - Si se produce un desacuerdo irreconciliable sobre dónde agrupar una determinada acción, puede duplicarse y colocarse en más de un grupo.

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.

- **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, etc. Cada una de estas líneas estratégicas contienen actividades concretas que vamos a poner en marcha. Todas ellas configuran la intervención a realizar para mejorar.
- **Describir el resultado** de forma clara y ordenada, para que todo el grupo y el resto del centro sepa 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. La Figura 12.2 es un ejemplo de diagrama de afinidades, resultado de una reflexión grupal sobre qué hacer para mejorar la detección, diagnostico y estudio inicial del hipertenso.

FIGURA 12.2. Ejemplo de diagrama de afinidades

¿QUÉ HACER PARA MEJORAR LA DETECCIÓN, DIAGNOSTICO 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 diagnostico inicial.
- Establecer algoritmo de procedimientos a realizar para incluir al paciente en programa.

Las *ventajas* de explicitar las acciones a tomar con un diagrama de afinidades son varias. En primer lugar implica 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. Adicionalmente 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 todos 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 a continuación.

El diagrama de afinidades produce 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

4. 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, como ya indicamos en la UT 6, 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 como el *diagrama de problemas anticipados* y el *diagrama de Gantt*; herramientas ambas que

pasamos a describir pero que, en nuestra experiencia, no llegan a hacerse necesarias en las propuestas de mejora habituales, sobre todo al principio de los programas o actividades de gestión de la calidad. La Tabla 12.4 contiene la relación de métodos que vamos a revisar.

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

TABLA 12.4. Métodos para el seguimiento del proyecto de mejora

- ANÁLISIS Y VIGILANCIA DE PROBLEMAS ANTICIPADOS
- DIAGRAMA DE GANTT
- CARTELES NARRATIVOS

4.1. 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 se presenta en la Figura 12.3. Como puede verse, en la parte izquierda del diagrama están las diversas acciones o actividades a realizar ordenada en secuencia temporal lógica, mientras que en el encabezamiento (o a veces en la parte de abajo) 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, pero que no vamos a considerar aquí por su mayor complejidad y una menor aplicabilidad en los ciclos de mejora normales.

FIGURA 12.3. Esquema para elaboración de un diagrama de gantt

	Unidades de Tiempo (días, semanas, meses, etc.)									
Tareas, Actividades	1	2	3	4	5	6	7	8	9	10
1.										
2.										
3.										
4.										
5.										
6.										
7.										
8.										
9.										

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.

En la Figura 12.4 puede verse un ejemplo simple de diagrama de Gantt, aplicado a la planificación operativa de los cambios a introducir en lo sistemas de registro, una de las líneas estratégicas acordadas en nuestro diagrama de afinidades.

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

FIGURA 12.4. Ejemplo de 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										

4.2. 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 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

La Tabla 12.5 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 última, más recomendable y más sencilla de las herramientas que sugerimos utilizar y que ya fue vista en la UT 6, puesto que lo más indicado es empezar su construcción a la vez que se inicia el Ciclo de Mejora.

TABLA 12.5. ESQUEMA PARA UN DIAGRAMA DE PROBLEMAS ANTICIPADOS

Resumen de la estrategia a implementar:						
¿Qué puede ir mal?	¿Qué probabilidad hay de que ocurra? (a)	¿Cómo es de importante 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: poca; 2: moderada; 3: mucha; 4: muchísima.

4.3. CARTELES NARRATIVOS DEL PROYECTO DE MEJORA

Como ya vimos en la UT 6 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 inicialmente 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.

5. SEGUIMIENTO DE LA IMPLEMENTACIÓN DE LA INTERVENCIÓN PARA MEJORAR

Los métodos y herramientas que acabamos de ver tienen como uno de sus principales objetivos facilitar la continuación y el seguimiento del proyecto de mejora. 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 discusiones 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 que 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 prevista o no explicitada en la estrategia diseñada, no hay que tener ningún reparo en ponerla en práctica. Si funciona, es decir si conseguimos finalmente mejorar, hemos hecho bien.

Cualquier actividad de seguimiento, formal o informal, es útil y conveniente en la medida que logre que la mejora se consiga.

BIBLIOGRAFÍA

- Leebow W, Ersoz CJ. The health care manager´s guide to continuous quality improvement. Chicago: AHA; 1991.
- Goldfield N, Pine M, Pine J. Measuring and managing health care quality. Maryland: Aspen; 1995.

13

REEVALUACIÓN, ANÁLISIS Y PRESENTACIÓN DE RESULTADOS COMPARATIVOS DE DOS EVALUACIONES

EMCA

Gestión de la Calidad Asistencial

CONTENIDO GENERAL

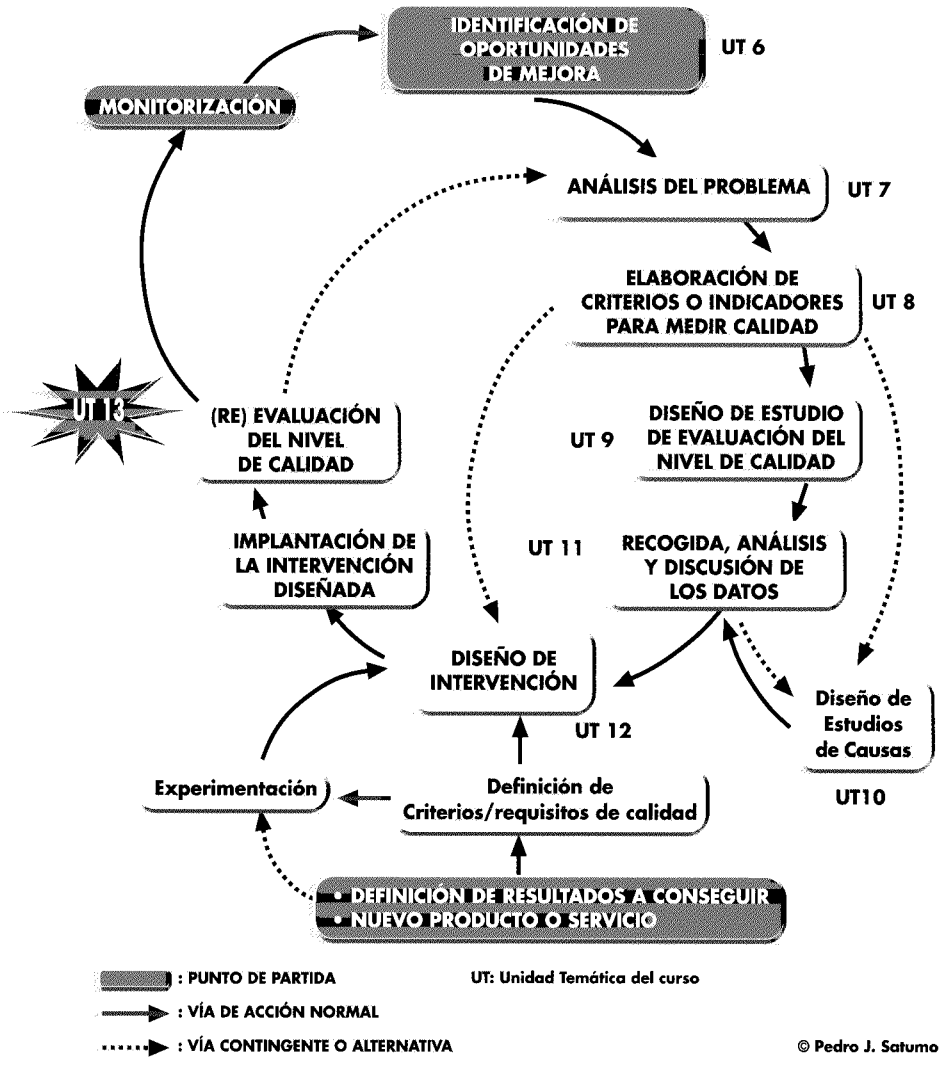
Una vez haya dado tiempo a que la intervención haya tenido efecto, es precisa una nueva medición de la calidad para documentar la mejora. En esta UT se detalla la metodología necesaria para cumplir con este objetivo. Se parte del hecho de que en la reevaluación lo que pretendemos no es solamente medir el nuevo nivel de calidad, sino también comprobar la hipótesis de mejora. Para este fin se explican las pruebas estadísticas aplicables y el análisis de los datos con un diagrama de Pareto antes-después. Por otra parte, se revisan, adaptándolos a la reevaluación, muchos de los conceptos de las UT 9 Y 11

ÍNDICE DE CONTENIDOS

1. Introducción
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. Análisis Gráfico de las diferencias.
5 ¿Qué hacemos a continuación? Cursos de acción tras la reevaluación.

OBJETIVOS ESPECÍFICOS

1. Diseñar la recogida de datos utilizando un esquema similar al empleado en la UT 11.
2. Analizar los datos en relación al nivel de calidad en esta evaluación, de forma similar a lo realizado con los datos de la primera evaluación (UT 11)
3. Explicar el concepto de significación estadística al comprobar la hipótesis de mejora.
4. Aplicar correctamente diversas pruebas de significación estadística para el contraste de la hipótesis de mejora.
5. Interpretar los resultados de estas pruebas para responder a la respuesta ¿Hemos mejorado?
6. Calcular la mejora relativa de cada criterio o requisito.
7. Presentar gráficamente los resultados comparativos del nivel de calidad de las dos evaluaciones con un gráfico "estrella" o " radar".
8. Construir un gráfico de Pareto antes-después.
9. 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.
10. Decidir cuál es el curso de acción más conveniente, a la vista de los resultados de la reevaluación.



"La vida sólo se comprende mirando hacia atrás, pero se debe vivir hacia delante"

Kierkegaard

1. INTRODUCCIÓN

En esta UT se abordan los aspectos metodológicos relacionados con la comparación entre dos evaluaciones y las consideraciones sobre la forma de continuar mejorando

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; (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 esta UT. En ella se contemplan las diferencias posibles y no deseables en los diseños de la primera y segunda

evaluación, la forma de cuantificar la diferencia en los niveles de calidad entre las dos, y su significación estadística, y la manera de representar esta diferencia gráficamente, tanto si queremos enfatizar el aumento de nivel de calidad criterio a criterio, como si nuestro interés primordial es analizar lo mejorado y lo que queda por mejorar.

2. DISEÑO DE LA REEVALUACIÓN. COMPONENTES Y DIFERENCIAS CON EL DISEÑO DE LA PRIMERA EVALUACIÓN

El diseño 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 la Tabla 13.1) y responden a la misma descripción que vimos en la UT 9, que es conveniente revisar a este respecto.

TABLA 13.1. Componentes de una evaluación

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
   - 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.

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 EN LA SEGUNDA EVALUACIÓN

A efectos de comparabilidad con la primera evaluación no es en absoluto deseable cambiar 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.

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

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 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 diagnostico), 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 como 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.

- **Criterios con periodo de tiempo único**. Tomemos el 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 ; también sería posible mantener el mismo marco muestral (todos los hipertensos diagnosticados hasta el momento de la segunda evaluación) pero en este caso el efecto de la intervención puede quedar diluido, aparte de que tendríamos que decidir qué hacer

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.

con los diagnosticados incorrectamente antes de la primera evaluación: excluirlos de la segunda evaluación o iniciar de nuevo el proceso diagnostico 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 una "segunda oportunidad" no contemplada en la primera evaluación. A efectos comparativos es más importante mantener la definición de la unidad de estudio, aunque cambie la definición del marco muestral o universo para la extracción de los casos a evaluar, o tengamos 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.

- **Criterio con periodo de tiempo muy extenso**. 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 o evaluar sólo los casos que no la cumplían 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, como en el anterior, 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 tanto si tiene como si no 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.

Cuando los criterios se refieren a periodos de tiempo en el proceso de atención que no son únicos (como el diagnostico) 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.

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.

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, y que tenga, como para la primera evaluación, un tamaño que permita estimar el nivel de calidad con una cierta precisión. Nada más.

En la UT 9 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.

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 o 0,10) necesitamos muestras más grandes, sobre todo si el cumplimiento en la primera evaluación fue en torno a 50%

Para tener una idea de en qué medida necesitamos o no muestras más grandes para detectar según qué diferencias, en la Tabla 13.2 figura el tamaño de la muestra en ambas evaluaciones 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 de la Tabla 13.2 serán explicados más adelante en esta UT. 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 ó 0,1 cuando las proporciones de cumplimiento en la primera evaluación son más extremas (0,05, 0,1 ó 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.

TABLA 13.2. Tamaño de muestra (n) necesario en cada evaluacion para detectar como significativas (p≤0,05, una cola) diversos niveles de diferencias entre dos evaluaciones (d), según la proporción de cumplimiento de la primera evaluación (p_1)

	PROPORCIÓN DE CUMPLIMIENTO EN LA PRIMERA EVALUACIÓN (P1)																		
	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**
d	150	237	313	377	432	475	507	529	540	540	529	507	475	432	377	313	237	150	53
0,05	69	102	130	154	174	190	201	208	211	210	205	195	181	163	140	114	83	48	32
0,08	49	69	87	101	114	123	130	134	135	134	130	123	114	101	87	69	49	*	*
0,10	33	44	54	62	69	74	78	80	80	79	76	71	65	58	48	37	*	*	*
0,13	*	35	42	48	53	56	59	60	60	59	56	53	48	42	35	*	*	*	*
0,15	*	*	30	34	37	40	41	42	41	40	38	36	32	*	*	*	*	*	*
0,18	*	*	*	*	31	32	33	34	33	32	31	*	*	*	*	*	*	*	*
0,20	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
0,23																			
0,25	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

(*): n< 30
d: diferencia entre las dos evaluaciones: p2 -p1

3. PRESENTACIÓN Y ANÁLISIS DE LOS DATOS: ESTIMACIÓN DEL NIVEL DE CALIDAD Y MEJORA CONSEGUIDA TRAS LA INTERVENCIÓN PARA MEJORAR

En la reevaluación estamos interesados en saber el nivel de calidad alcanzado pero también, y sobre todo, en documentar la mejora conseguida; es decir: en comparar adecuadamente los resultados de las dos evaluaciones

Una vez resueltas las dudas sobre el diseño de la reevaluación, efectuamos la nueva recogida de datos y nos encontramos ante una situación semejante a la de la primera evaluación en cuanto a las estimaciones o inferencias que podemos hacer con la muestra recogida. Los métodos estadísticos en este sentido van a ser idénticos a los que vimos para 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 del nivel de calidad (puntual e intervalo), en ambas evaluaciones. En consecuencia vamos a realizar una estimación puntual de la diferencia y *además* deberemos 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 ellos, siguiendo el resumen que figura en la Tabla 13.3 y ejemplificando los cálculos con los datos de la Tabla 13.4.

3.1. ESTIMACIÓN DEL NIVEL DE CALIDAD

El objetivo y los procedimientos son idénticos a los que vimos en la UT 11. 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 la Tabla 13.3, pero existe una explicación más extensa en la UT 11, a la cual debemos remitirnos en caso de duda en este punto. La Tabla 13.4 contiene ambas estimaciones, puntual y de intervalo, para los criterios considerados en dos evaluaciones.

TABLA 13.3. Análisis de los datos de una reevaluación. Resumen de componentes y métodos.

1 ESTIMACIÓN DEL NIVEL DE CALIDAD*

- Estimación puntual: $p_2 = \frac{n^\circ \text{ de cumplimientos}}{\text{tamaño de la muestra}}$
- Intervalo de confianza (95%): $p_2 \pm 1{,}96\sqrt{\frac{p_2\,(1-p_2)}{n}}$

2 ESTIMACIÓN DE LA MEJORA CONSEGUIDA

- Mejora absoluta: $p_2 - p_1$
- Mejora relativa: $\frac{p_2 - p_1}{1 - p_1}$

3 SIGNIFICACIÓN ESTADÍSTICA DE LA MEJORA

- Calcular valor de z para la diferencia entre las dos evaluaciones(**):
- Diferencia entre proporciones: $z = \frac{p_2 - p_1}{\sqrt{p\,(1-p)\,(\frac{1}{n_1} + \frac{1}{n_2})}}$
- Diferencia entre medias: $z = \frac{\bar{\chi}_2 - \bar{\chi}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Ver significación en tablas de la distribución normal (una cola)

*: Procedimiento idéntico al de la primera evaluación. Ver las fórmulas para medias e intervalos de confianza según los diferentes métodos de muestreo.

**: Condiciones de aplicación:

- para proporciones: ambos $n_1.p$ y $n_2.p$ deben ser ≥5. Alternativa: Test exacto de Fisher.

- para medias: ambos n_1 y n_2 deben ser >30. Alternativa: Test no paramétrico ó t de Student

p_2: proporción de cumplimientos en la reevaluación; p_1: proporción de cumplimientos en la primera evaluación.

n_2: muestra en la reevaluación; n_1: muestra en la primera evaluación.

p: proporción de cumplimientos conjunta (primera y segunda evaluación juntas) =

$$\frac{n^\circ \text{ cumplimientos en la } 1^a + n^\circ \text{ cumplimientos en la } 2^a}{n_1 + n_2}$$

$\bar{\chi}_2$: media en la reevaluación; $\bar{\chi}_1$: media en la primera evaluación.

s_2: desviación estándar de los datos de la muestra en la reevaluación; s_1: desviación estándar de los datos de la muestra en la primera evaluación.

Las estimaciones puntual y de intervalo del nivel de calidad se realizan de forma idéntica a las de la primera evaluación

TABLA 13.4. Evaluacion 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

	1ª EVALUACIÓN	2ª EVALUACIÓN
CRITERIO	%CUMPLIMIENTO (IC 95%)	%CUMPLIMIENTO (IC 95%)
1. Diagnostico 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 carotideos	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)

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 de la Tabla 13.4 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 75-25). De igual manera podemos calcular la mejora absoluta para cada uno de los criterios de la Tabla 13.4, 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 (ó 100%, máximo cumplimiento posible) menos el nivel de cumplimiento en la primera evaluación (en proporción o porcentaje, según sea 1 ó 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): $\frac{0,95-0,75}{1-0,75} = \frac{0,2}{0,25} = 0,8$ o 80%

- Criterio 3 (Antecedentes personales): $\frac{0,98-0,90}{1-0,90} = \frac{0,08}{0,1} = 0,8$ o 80%

- Criterio 5 (Consumo alcohol): $\frac{0,50-0,10}{1-0,10} = \frac{0,4}{0,9} = 0,44$ o 44%

- Criterio 10 (Soplos carotideos): $\frac{0,75-0,55}{1-0,55} = \frac{0,20}{0,45} = 0,44$ o 44%

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).

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.

3.3. ¿HEMOS MEJORADO? SIGNIFICACIÓN ESTADÍSTICA DE LA DIFERENCIA ENTRE LAS DOS EVALUACIONES

Hay que realizar algunos cálculos estadísticos para poder concluir si la diferencia encontrada es real o puede ser explicada por los errores del muestreo

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*, 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 la UT 11 para el intervalo de confianza del nivel de cumplimiento. Responde a la fórmula general:

parámetro estimado ± z * desviación estándar del parámetro estimado

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

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 p2=0,95 y p1=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 exac-

tamente 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.

Para concluir si hemos mejorado más que calcular el intervalo de confianza de la diferencia, conviene utilizar un test estadístico apropiado

a) ¿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 seria 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 (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).

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 α, error tipo I y valor p

Pero ¡ojo!: 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 α 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 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 ≤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*.

b) 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 la Tabla 13.3):

$$z = \frac{p_2 - p_1}{\sqrt{p\,(1-p)\,(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Entre los tests estadísticos que se pueden utilizar, seleccionamos el cálculo del valor z (valor estadístico en la distribución normal) para la diferencia encontrada

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.

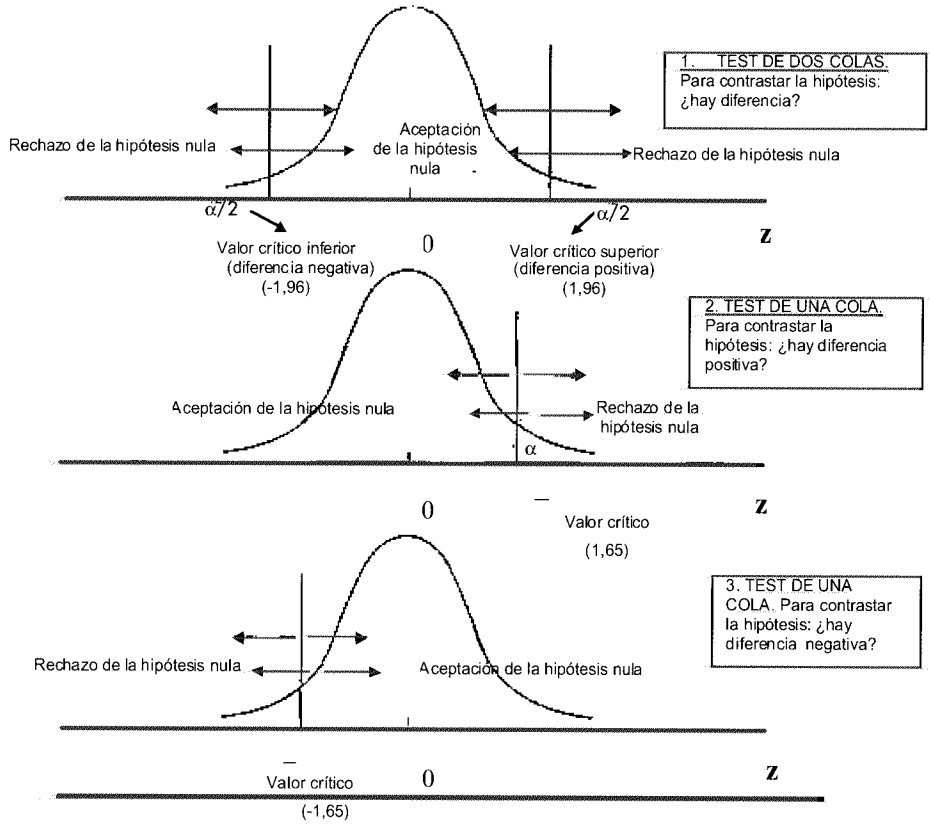
Como es evidente al estar p_2-p_1 en el numerador, si no hay diferencia entre las dos evaluaciones el valor de *z* es cero, y obtendremos valores de *z* tanto mayores cuanto mayor sea la diferencia encontrada. La cuestión es ¿cuánto puede llegar a ser de grande el valor de *z* y seguir siendo compatible con una situación 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 ≤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*.

c) 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 (Figura 13.1), 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.

FIGURA 13.1. Test de hipótesis de la diferencia entre dos evaluaciones. significación (probabilidades) de una o dos colas según la formulación de la hipótesis alternativa*



* La hipótesis nula es en todos los casos que no hay diferencia. Los valores críticos están establecidos para una significación de 0,05 (5%).

Fte.: Adaptado de Richards LD, La Cava JJ.

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

¿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 segundo 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 de la Tabla 13.4.

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:

$$z = \frac{p_2 - p_1}{\sqrt{p\,(1-p)\,(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0,95 - 0,75}{\sqrt{(0,85)\,(0,15)\,(\frac{1}{60} + \frac{1}{60})}} = 3,07$$

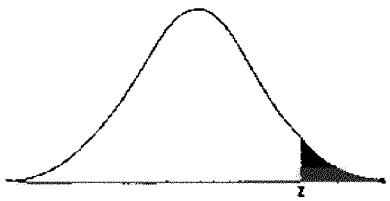
Para ver el nivel de significación buscamos en la tabla de la distribución normal (una cola), que reproducimos como Tabla 13.5, la probabilidad que corresponde a ese valor de *z*. Para ello buscamos en la primera columna el valor de las unidades y primer decimal (3,0) y en la fila del encabezamiento 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.

Para el criterio 3, p_2=0,98; p_1=0,9; $p = \frac{0,98 + 0,9}{2} = 0,94$

$$z = \frac{0,98 + 0,9}{\sqrt{(0,4)\,(0,06)\,(\frac{1}{60} + \frac{1}{60})}} = 1m85$$

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.

TABLA 13.5. Probabilidades en un extremo o cola de la curva estándar normal



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	10,033	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

Fte: Adaptado de Colton T.

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, 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 la Tabla 13.5, valores de z ≥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.

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

La Tabla 13.6 contiene el análisis estadístico completo de los criterios de la Tabla 13.4, 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.

TABLA 13.6. Evaluación de la calidad del diagnostico, 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	Significación estadística *p*
1. Diagnostico 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)	—	—	—
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	42,5%	0,031
9. Pulsos periféricos	50,0 (±12,7)	60,0 (±12,4)	10	20%	NS
10. Soplos carotideos	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

p_1= cumplimiento en la primera evaluación.
p_2= cumplimiento en la segunda evaluación.
NS: No significativa (p >0,05)

4. ANÁLISIS GRÁFICO DE LAS DIFERENCIAS

4.1. REPRESENTACIÓN GRÁFICA DE LAS DIFERENCIAS EN LOS NIVELES DE CALIDAD ENTRE DOS EVALUACIONES

La representación gráfica de las diferencias del nivel de calidad completa el análisis comparativo de las dos evaluaciones

Al igual que veíamos en la UT 9 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 la UT 9, solo que añadiendo los resultados de la segunda evaluación en la forma que resulte más útil e informativa.

En la Figura 13.2 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 de la Tabla 13.4.

4.2. 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 segunda evaluación; a lo que teníamos en la primera 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 de la Tabla 13.4 en el gráfico de la Figura 13.3.

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 la UT 11, que pierde en claridad, sobre todo en la comparación criterio a criterio, a medida que aumenta el número de criterios representados. La Figura 13.4 es el gráfico de estrella comparando las dos evaluaciones de la Tabla 13.4.

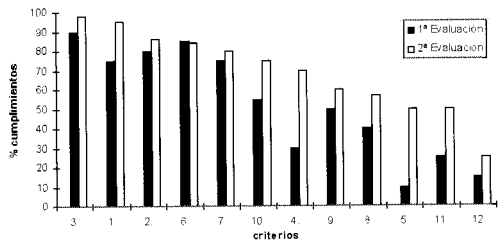
No es frecuente analizar y representar las diferencias en intervalos de confianza. Lo más normal es representar las estimaciones puntuales (barras, estrella) o defectos (Pareto) e indicar, si acaso, en qué criterio es la diferencia estadísticamente significativa. No obstante, si quisiésemos plasmar gráficamente las diferencias en las estimaciones de intervalos de confianza, el gráfico de elección sería el de tipo *box and whiskers*, cuya explicación vimos también en la UT 11.

FIGURA 13.2. Representación y análisis gráfico de la comparación entre dos evaluaciones. Resumen de objetivos y esquemas de alternativas

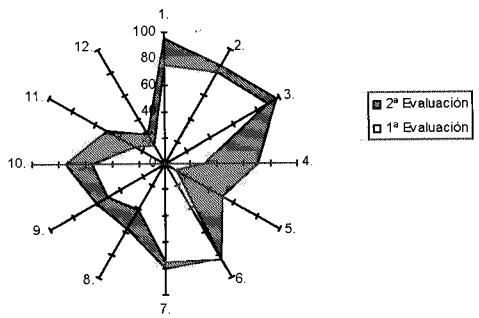
1. REPRESENTACIÓN DE DIFERENCIAS EN EL NIVEL DE CALIDAD

• DIFERENCIAS EN ESTIMACIONES PUNTUALES

— Diagrama de barras comparativas

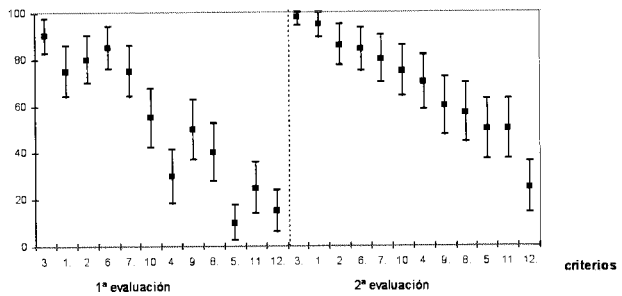


— Gráfico de estrella.



• DIFERENCIAS CON INTERVALOS DE CONFIANZA

— Gráfico box and whiskers comparativo



2. ANÁLISIS DE LA MEJORA Y DEFECTOS POR MEJORAR

— Diagrama de Pareto antes-después

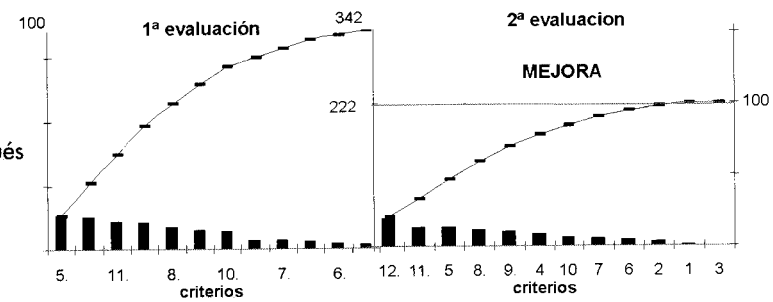
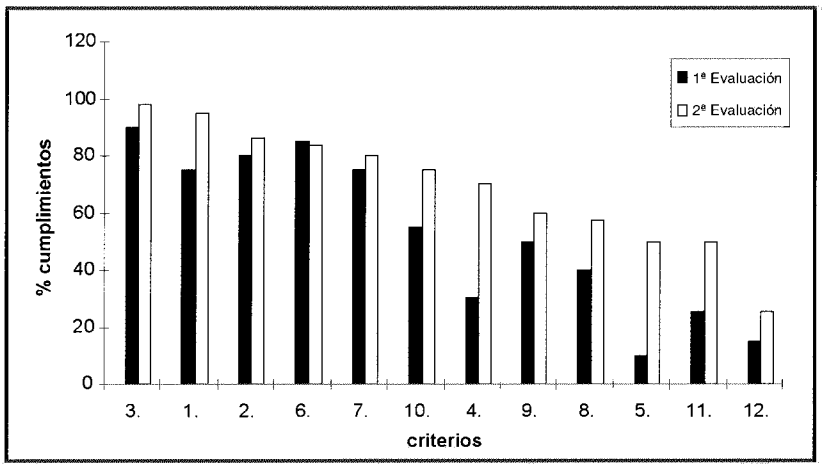


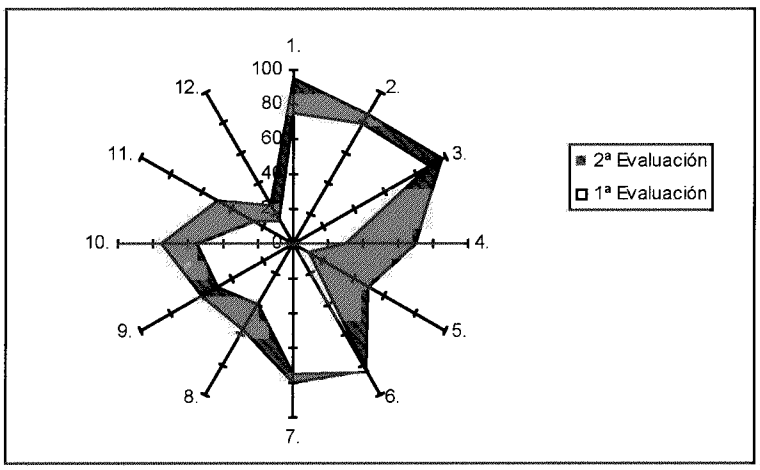
FIGURA 13.3. Comparación gráfica de dos evaluaciones. Gráfico de barras



CRITERIOS

1. Diagnostico 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 carotideos
11. Edemas
12. Fondo de ojo

FIGURA 13.4. Comparación gráfica de dos evaluaciones. Gráfico de estrella

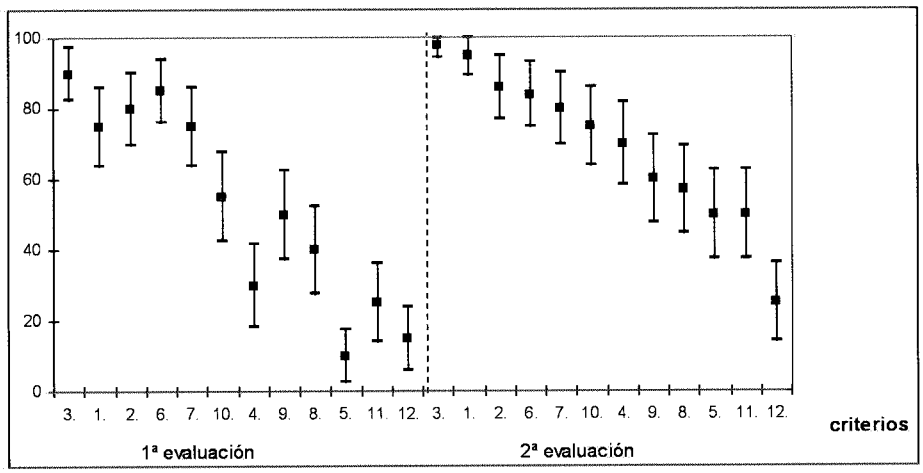


CRITERIOS

1. Diagnostico 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 carotideos
11. Edemas
12. Fondo de ojo

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 13.5 se han representado los datos de la Tabla 13.4 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).

FIGURA 13.5. Comparación grafica de dos evaluaciones. Box and whiskers antes-después



CRITERIOS

1. Diagnostico 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 carotideos
11. Edemas
12. Fondo de ojo

La representación gráfica más completa e informativa es el gráfico de Pareto antes-después

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 gráfico, y que se convierte así en la representación gráfica de elección, caso de querer seleccionar sólo una.

4.3. ANÁLISIS GRÁFICO DE LA MEJORA Y DEFECTOS POR MEJORAR: DIAGRAMA DE PARETO ANTES-DESPUÉS

En la UT 11 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.

Para que el Pareto antes-después resulte más informativo hay que mantener constante en los dos diagramas el eje del número absoluto de defectos.

En los manuales sobre métodos para el control y la mejora de la calidad se pueden encontrar varias formas de diagrama de Pareto 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 tan visualmente 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 13.2. 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 13.6 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 13.6c), o mantener el orden de izquierda a derecha para los

dos (Figura 13.6). 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 de la Tabla 13.4, convertidos en información útil para los diagramas de Pareto son los que figuran en la Tabla 13.7. Con ellos hemos construido los tres diagramas antes-después de la Figura 13.6, 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.

TABLA 13.7. Comparación de dos evaluaciones, frecuencia absoluta, relativa y acumulada de los incumplimientos de los criterios evaluados

1ª EVALUACIÓN				2ª EVALUACIÓN			
CRITERIO	Nº INCUM (Frecuencia absoluta)	% (Frecuencia relativa)	Frecuencia acumulada	CRITERIO	Nº INCUM (Frecuencia absoluta)	% (Frecuencia relativa)	Frecuencia acumulada
5.Consumo 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 carotideos	27	7,9	83,3	**10.**Soplos carotideos	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**	

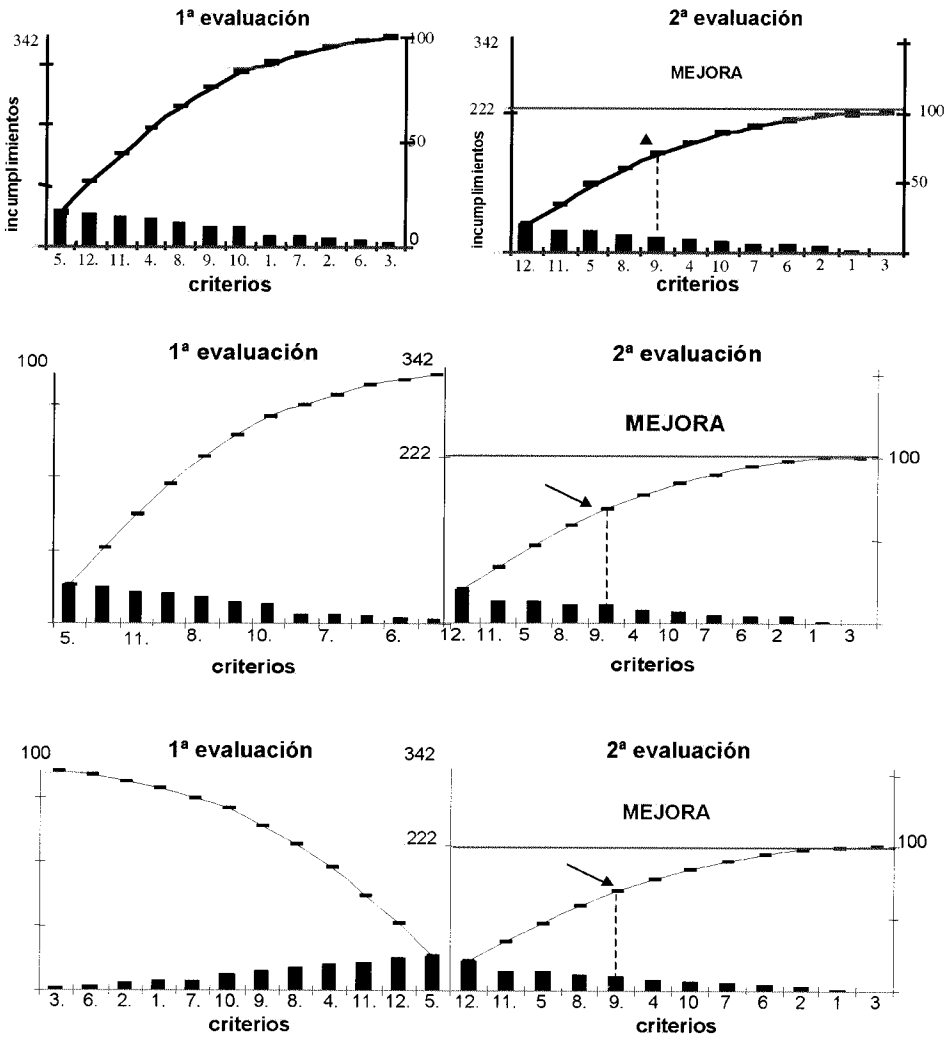
En caso de que la muestra empleada en la segunda evaluación no sea igual que la de la primera hay que hacer un pequeño ajuste para poder utilizar cualquiera de las tres modalidades de Pareto antes-después que hemos propuesto. El ajuste consiste en calcular, por simple regla de tres, lo que supondría 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 la Tabla 13.7 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 evaluaciones; 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 segunda evaluación si el tamaño de muestra fuese como el de la primera evaluación. De la siguiente forma:

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

$$\frac{\text{defectos en la } 2^{a} \text{ evaluación } (d_2)}{\textit{muestra} \text{ en la } 2^{a} \text{ evaluación } (n_2)} = \frac{x}{\text{muestra en la } 1^{a} \text{ evaluación } (n_1)}$$

De donde, $x = \dfrac{d_2 \,?\, n_1}{n_2}$

FIGURA 13.6. Comparación grafica de dos evaluaciones: gráfico de Pareto antes-después



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 criterio 12 en Pareto comparativo de la 2ª evaluación, si n_2 fuese 80 (y no 60 como n_1) sería:

$$x = \frac{45 \cdot 50}{80} = 34$$

de forma que los 45 incumplimientos de la muestra de 80 casos contarían como 34 para la comparación con la primera 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.

5. ¿QUÉ HACEMOS A CONTINUACIÓN? CURSOS DE ACCIÓN TRAS LA REEVALUACIÓN

En general, dentro de nuestras actividades para mejorar la calidad de la atención que ofrecemos a nuestra población, siempre tendremos la opción de iniciar nuevos ciclos de mejora, diseñar mejor los servicios que ofrecemos, y construir y monitorizar indicadores relevantes y representativos de los aspectos que más nos interesen de nuestro trabajo.

Sin embargo, tendremos que decidir en primer lugar qué hacemos en relación con el ciclo de mejora cuya reevaluación acabamos de realizar. Obviamente, lo que decidamos hacer va a estar condicionado por el resultado obtenido en nuestro ciclo 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?..).

no hemos mejorado, hay que reiniciar el ciclo y pensar en qué podemos haber fallado

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 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.

BIBLIOGRAFÍA

- Richard LE, LaCava JJ. Bussiness statistics. New York: McGraw Hill; 1983.
- Lemeshow S, Hosmer DW, Klar J, Lwanga SK. Adequancy of sample size in health studies. New York: WHO/John Wiley and sons; 1992.
- Gitlow H, Gitlow S, Oppenheim A, Oppenheim R. Tools and methods for the improvement of quality. Boston: Irwin; 1989.

ACTIVIDADES PARA LA MONITORIZACIÓN. CONSTRUCCIÓN Y ANÁLISIS DE INDICADORES. PLANES DE MONITORIZACIÓN

EMCA

Gestión de la Calidad Asistencial

CONTENIDO GENERAL

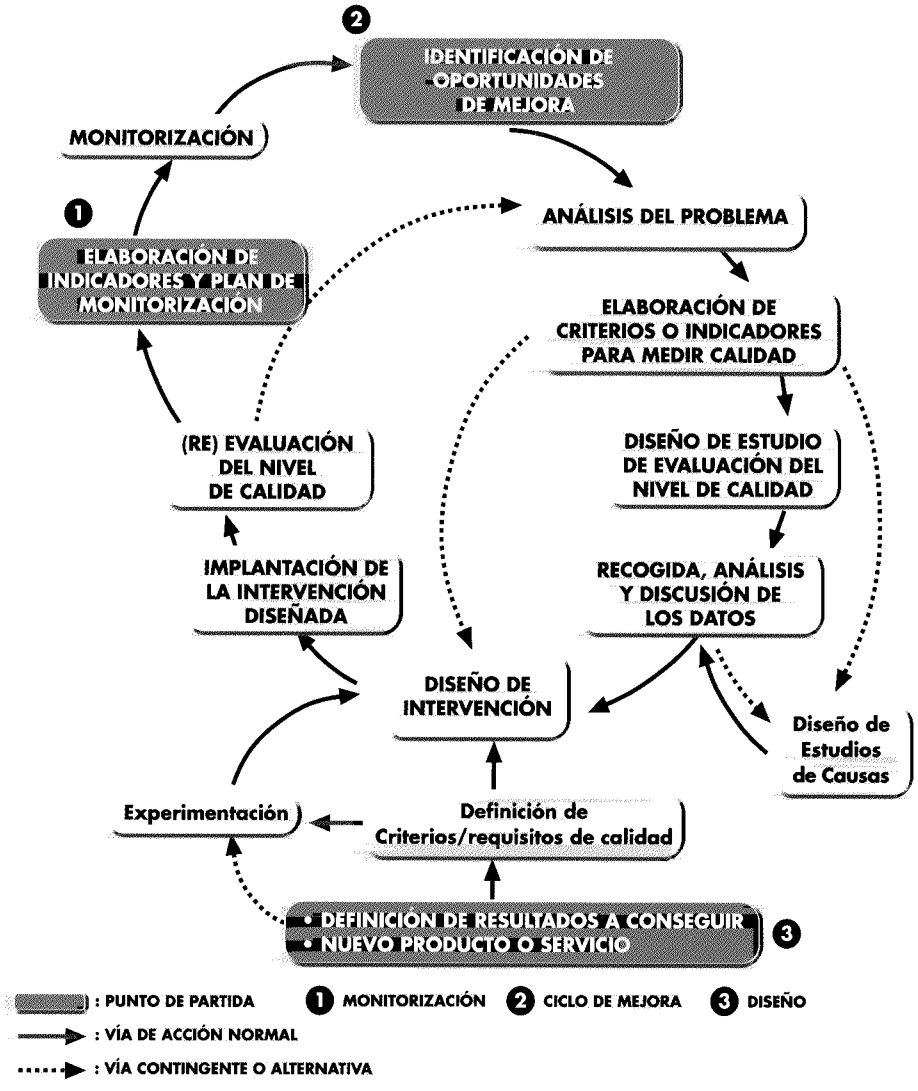
Esta UT ofrece una revisión de los principales conceptos y esquemas metodológicos de la monitorización, y las siguientes UT las herramientas y técnicas que los llevan a cabo. Con ello se completa la visión panorámica de este importante grupo de actividades.

ÍNDICE DE CONTENIDOS

1. Introducción.
2. Evaluar, diseñar y monitorizar la calidad: los tres grupos de actividades que llevan a la práctica el compromiso con la calidad.
3. Medición de la calidad. Situaciones y objetivos.
4. Monitorización de la calidad: ¿qué incluye?
5. Puntos de partida para monitorizar: origen de los indicadores.
6. Construcción y análisis de indicadores. Los puntos clave.
7. Componentes de un Plan de Monitorización. Tipos de Plan.
8. Métodos de monitorización según las características del plan de monitorización.

OBJETIVOS ESPECÍFICOS

1. Distinguir las principales actividades incluidas en la monitorización o control de los niveles de calidad.
2. Explicar la relación entre Monitorización y Ciclos de Mejora.
3. Delimitar las diferencias prácticas entre criterio o requisito e indicador.
4. Definir los principales componentes de que consta un plan de monitorización.
5. Describir los objetivos generales de los planes de monitorización.
6. Clasificar los planes de monitorización en función de la periodicidad de las mediciones.
7. Enumerar diversas metodologías aplicables en la monitorización de indicadores.



"Ningún hombre de temperamento científico afirma que lo que ahora es creído en ciencia sea exactamente verdad; afirma que es una etapa en el camino hacia la verdad".

Bertrand Russel

1. INTRODUCCIÓN

Después de una introducción a los conceptos básicos para la gestión de la calidad y de haber seguido paso a paso, con todas sus implicaciones metodológicas un Ciclo de Mejora, en esta UT volvemos al principio para, desde ahí, adentrarnos en las vías de la monitorización.

Vamos a dar una visión general de lo que implica añadir plenamente las actividades de monitorización a los programas de gestión de la calidad. En concreto, vamos a subrayar la importancia de contar con indicadores válidos y de definir de forma correcta los planes de monitorización, incluyendo los métodos o esquemas de mediciones más oportunos y eficientes que serán desarrollados en las UT siguientes.

En esta Unidad Temática se da una visión panorámica de los métodos y actividades de monitorización de la Calidad.

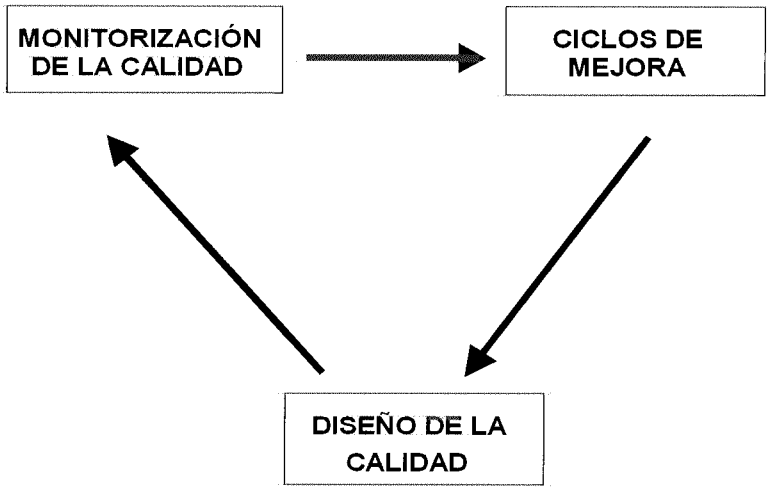
2. EVALUAR, DISEÑAR Y MONITORIZAR LA CALIDAD: LOS TRES GRUPOS DE ACTIVIDADES QUE LLEVAN A LA PRÁCTICA EL COMPROMISO CON LA CALIDAD

En la UT 3, vimos que se pueden definir tres grupos de actividades (ciclos de mejora, monitorización y diseño) en los programas de Gestión de la Calidad, y que el que se realicen unas u otras y sus combinaciones va a ser una de las características definitorias del programa. Posteriormente, en la UT 5 razonábamos el comenzar con un ciclo de mejora como estrategia para la implantación progresiva de las actividades del programa. La distinción entre los tres grupos de actividades la hacíamos en función de los diferentes puntos de partida y objetivos inmediatos que persiguen, según se resume en la tabla que reproducimos como Tabla 14.1. Por otra parte, indicábamos que los tres tipos de actividades se relacionan entre sí, y que si bien los ciclos de mejora tienen sentido (principio y final) por sí mismos, (e incluyen actividades de diseño, al definir la intervención para mejorar), tanto la monitorización como el diseño necesitan ser complementados con al menos otra de las actividades (la monitorización con ciclos de mejora, y el diseño con la monitorización). Esta interrelación se representa en la Figura 14.1.

TABLA 14.1. Grupos de actividades de los programas de Gestión de la Calidad. Puntos de partida normales y objetivos inmediatos

GRUPO DE ACTIVIDADES	PUNTO DE PARTIDA	OBJETIVO INMEDIATO
Ciclos de mejora	• Identificación de un problema de calidad u oportunidad de mejora en algún aspecto de los servicios que se ofrecen.	• Solucionar el problema. • Aprovechar la oportunidad de mejora descubierta.
Monitorización	• 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.	• Identificación de problemas de calidad u oportunidades de mejora.
Diseño	• Programación de un nuevo servicio a ofrecer. • Identificación de necesidades y expectativas de los usuarios. • Identificación de parámetros y resultados a conseguir.	• Diseñar los procesos de atención para conseguir los resultados deseados predeterminados.

FIGURA 14.1. Los grupos de actividades de los Programas de Gestión de la Calidad



Después de haber revisado y practicado los métodos para los ciclos de mejora, es conveniente aclarar un poco más en detalle lo que implica comenzar las actividades del programa con la monitorización. Comencemos por resumir y aclarar la relación entre ciclos de mejora y monitorización a través, sobre todo, de la comprensión de las situaciones y objetivos que podemos tener cuando medimos la calidad.

3. MEDICIÓN DE LA CALIDAD. SITUACIONES Y OBJETIVOS

En los programas de gestión de la calidad, ésta puede ser medida en tres principales situaciones o momentos (en la evaluación, en los ciclos de mejora y su reevaluación tras intervenir, y en la monitorización) que responden a dos objetivos: la estimación del nivel de calidad y/o la comparación con niveles preestablecidos. La distinción entre las diversas situaciones y objetivos es importante porque pueden requerir acercamientos estadísticos diferentes y determinar la utilidad o no de determinados métodos. Los principales objetivos, tal como se resume en la Tabla 14.2, son los siguientes:

- **Estimación del nivel de calidad.** Es decir, medición del cumplimiento de requisitos, criterios o indicadores de calidad del servicio o aspecto del servicio que queremos evaluar. Esta estimación puede hacerse para diferentes propósitos: por ejemplo, para saber si hay o no espacio para la mejora; al inicio de un ciclo de mejora para saber la situación de partida; para comparar diversos centros o servicios, etc. En cualquier caso, lo que queremos es saber, con una determinada precisión, a qué nivel concreto está el servicio o el aspecto del servicio cuya calidad medimos.

Al medir la calidad podemos estar interesados en estimar su nivel o en comparar el resultado de la medición con un nivel preestablecido.

- **Comparación de la situación actual con un nivel preestablecido.** Esta comparación puede hacerse en dos tipos de situaciones y para cumplir dos objetivos diferentes: (i) documentar la mejora conseguida tras la implementación de la intervención en un ciclo de mejora, en cuyo caso comparamos con la medición previa, de base; o (ii) asegurarnos que estamos a un nivel deseable

predeterminado (que podemos llamar estándar), con el cual nos comparamos.

TABLA 14.2. Medición de la calidad. Situaciones, objetivos y métodos

SITUACIÓN	OBJETIVOS	MÉTODOS
1. EVALUACIÓN EN CICLOS DE MEJORA.	ESTIMACIÓN DEL NIVEL DE CUMPLIMIENTO DE LOS CRITERIOS O REQUISITOS DE CALIDAD.	ESTADÍSTICA DESCRIPTIVA (PROPORCIONES, PORCENTAJES, MEDIAS Y SUS INTERVALOS DE CONFIANZA).
2. REEVALUACIÓN TRAS INTERVENIR.	COMPARAR CON LA SITUACIÓN DE PARTIDA (¿Hemos mejorado?).	COMPARAR PROPORCIONES O MEDIAS DE LAS DOS EVALUACIONES Y VER SIGNIFICACIÓN ESTADÍSTICA DE LA DIFERENCIA.
3. MONITORIZAR (MEDICIONES REPETIDAS Y PLANIFICADAS).	COMPROBAR SI ESTAMOS O NO A UNOS NIVELES DE REFERENCIA PREVIAMENTE ESTABLECIDOS.	• ESTIMACIONES SUCESIVAS DEL NIVEL DE CUMPLIMIENTO. • DECISIÓN CON BASE ESTADÍSTICA (ACEPTACIÓN/ RECHAZO) SIN ESTIMAR EL NIVEL REAL DE CUMPLIMIENTO.

Cuando queremos tener una estimación del cumplimiento de los criterios o requisitos, situación que se trata en la UT 11, utilizamos normalmente una muestra, que obtenemos de forma aleatoria con el método más apropiado al caso (aleatoria simple, sistemática, estratificada, etc.).

La comparación con niveles preestablecidos puede ser para documentar diferencias (mejoras) en relación a otras mediciones o para identificar problemas de calidad, comparando los resultados de la medición a un estándar preestablecido. En este segundo caso, algunos de los métodos desarrollados en la industria son de gran utilidad.

Cuando queremos comparar nuestra medición con un nivel preestablecido, las dos situaciones diferentes que hemos mencionado (documentar la mejora o controlar que estamos a unos niveles predeterminados) pueden utilizar métodos diferentes, aunque ambas tienen en común el que pueden ser consideradas como una comprobación de hipótesis previas: ¿hemos mejorado? y ¿estamos a unos niveles determinados preestablecidos?

Documentar la mejora conseguida (contenido central de la UT 13) es en realidad comparar dos estimaciones, antes y después de la intervención para mejorar. En esencia, es comprobar la hipótesis nula de ausencia de mejora, frente a la hipótesis alternativa de haber logrado mejorar; es decir, contestar a la pregunta ¿cuál es la probabilidad de encontrar, debido al azar, la diferencia que hemos encontrado entre las dos evaluaciones si no hay diferencia real? o, lo que es lo mismo, ¿es la mejora estadísticamente significativa? Para ello podemos utilizar, como vimos en la UT 13, los tests estadísticos usuales viendo significación de una cola. Ni en este caso, ni en ningún otro para el que nos basemos en estimaciones del nivel de calidad, existen métodos especiales desarrollados en el campo del control de la calidad en la industria.

- **Comparar la situación actual a un estándar prefijado.** Sin embargo, los métodos desarrollados en la industria son muy útiles y pueden ser adaptados con fortuna a la gestión de la calidad en los servicios de salud, si lo que queremos es *comparar la situación actual a un estándar prefijado*, con el fin de saber si estamos o no ante un problema de calidad que merece ser investigado o sujeto a intervención para mejorar. Éste es precisamente el objetivo principal de lo que hemos llamado monitorización de la calidad. Los indicadores y los estándares pueden tener origen diverso, pero los métodos de monitorización van a estar condicionados fundamentalmente por la forma en que queramos cubrir el objetivo de la identificación de problemas y la frecuencia de las mediciones, tal como veremos a continuación.

4. MONITORIZACIÓN DE LA CALIDAD: ¿QUÉ INCLUYE?

Entendemos por monitorización la *medición sistemática y planificada de indicadores de calidad,* una actividad conducente a controlar que estamos a unos niveles preestablecidos y que tiene como objetivo identificar la existencia o no de situaciones problemáticas que hay que evaluar o sobre las que hay que intervenir. Los indicadores a monitorizar pueden derivarse de ciclos de mejora, de actividades de diseño de los servicios, o ser fruto de una selección de aspectos o servicios relevantes de nuestro centro cuya calidad nos interesa controlar. En cualquier caso, deben ir acompañados de un plan de monitorización que incluya periodicidad, mecanismos para la recogida de datos y métodos de interpretación de los mismos.

La monitorización es la medición sistemática, repetida y planificada de indicadores de calidad, con el fin de identificar situaciones problemáticas.

En resumen, los dos componentes metodológicos básicos de la monitorización son: (i) La identificación, selección o *construcción de los indicadores* a medir; y (ii) la definición del *plan de monitorización*, incluyendo como mínimo, tal como veremos más adelante, la periodicidad de las mediciones y el método con que se van a realizar.

Monitorizar implica identificar los indicadores a medir y definir el plan de monitorización.

5. PUNTOS DE PARTIDA PARA MONITORIZAR: ORIGEN DE LOS INDICADORES

A la monitorización puede llegarse esencialmente a través de tres situaciones diferentes resumidas en la Tabla 14.3: (i) tras la elaboración y adopción de un listado de indicadores que representen los aspectos o servicios más relevantes que ofrecemos; (ii) tras completar un ciclo de mejora y seleccionar indicadores sobre el aspecto o servicio mejorado; y (iii) tras implantar un nuevo diseño de la forma de ofrecer un servicio reformado o añadido a nuestra actividad.

TABLA 14.3. Monitorización de la calidad. Situaciones y objetivos

SITUACIÓN	OBJETIVOS
• Medición de listado de indicadores sobre aspectos o servicios relevantes.	• Asegurarse que están a niveles aceptables. • Identificar aspectos-problema, o priorizar oportunidades de mejora.
• Medición de indicadores sobre un aspecto o servicio que ha sido sometido a ciclo de mejora.	• Asegurarse que se mantiene la mejora conseguida.
• Medición de indicadores sobre un aspecto o servicio de nuevo diseño.	• Asegurarse que el diseño funciona como estaba previsto.

La monitorización de un listado de indicadores sobre la actividad del centro es la situación más común como punto de partida para monitorizar, pudiendo servir de entrada a las actividades de gestión de la calidad, tal como contemplan y recomiendan organizaciones como la Joint Commission y diversos autores. Con algunas matizaciones, es también la base de programas externos como la valoración de las llamadas Normas Técnicas de la Cartera de Servicios del INSALUD, o la evaluación de los diversos ítems para los programas de acreditación de centros. El objetivo de estas mediciones es asegurarse que estos indicadores están a niveles aceptables, y, a la vez, identificar aspectos-problema que hay que evaluar y mejorar.

Una situación diferente es la medición de indicadores sobre un aspecto o servicio que ha sido sometido a un ciclo de mejora. En este caso, si bien la actividad de base es la misma (construcción de indicadores, normalmente seleccionando criterios de los utilizados en la evaluación) el objetivo es asegurarse que se mantiene la mejora conseguida, al menos durante un tiempo que nos parezca prudencial en función del aspecto que estemos monitorizando. Sin embargo, el objetivo en último término es, como en el caso de los listados de indicadores, identificar situaciones en las que (en este caso por no haberse mantenido la mejora conseguida) es necesario investigar (de nuevo) las causas de la calidad deficiente.

A la monitorización se puede llegar tras un ciclo de mejora, tras la elaboración de un listado de indicadores sobre la actividad del centro, o tras actividades de diseño de la calidad

Una situación semejante, en cuanto se parte de la implantación de una intervención diseñada para un determinado aspecto o servicio, es la elaboración y medición de indicadores sobre un aspecto o servicio de nuevo diseño. En este caso lo que nos interesa es fundamentalmente asegurarnos que el diseño funciona como estaba previsto, pero en último término, al igual que en las otras situaciones, se monitoriza para identificar la existencia o no de problemas que merezcan ser investigados.

En las tres situaciones, por lo tanto, el objetivo último de la monitorización es la identificación de problemas y los métodos de medición pueden ser semejantes en función de cómo establezcamos el plan de mediciones, con independencia de la situación que da origen a los indicadores y a la medición en sí. Las diferencias en la práctica entre medir para evaluar (para ciclos de mejora) y monitorizar se resumen en la Tabla 14.4.

TABLA 14.4. Diferencias entre evaluación y monitorización de la calidad

EVALUACIÓN	MONITORIZACIÓN
COMPRENSIÓN AMPLIA DE PROCESO	MENOS AMPLIA/RESUMEN
VARIOS ASPECTOS/CRITERIOS	POCOS ASPECTOS/UN CRITERIO
ADHOC, OCASIONAL	CONTINUA/RUTINARIA
CONDUCE A INTERVENCIÓN Y MONITORIZACIÓN	CONDUCE A EVALUACIÓN (si no se alcanza el umbral o aparece el criterio centinela)

Como ya queda dicho, sea cuál sea el punto de partida, las dos principales actividades que se necesitan para la monitorización son la construcción y/o análisis de los indicadores y el diseño del plan de monitorización.

6. CONSTRUCCIÓN Y ANÁLISIS DE INDICADORES. LOS PUNTOS CLAVE

Un *indicador* es un aspecto relevante que resume en la medida de lo posible la calidad de la actividad o problema que se desea monitorizar. Su utilización es más que otra cosa como herramienta de *screening* para detectar problemas que hay que evaluar, sea como método en sí de identificación de problemas o como mecanismo de asegurarnos que nos mantenemos a niveles logrados tras ciclos de mejora.

Sea cual sea el origen de la monitorización, ya visto anteriormente, lo que hay que monitorizar son buenos indicadores, es decir indicadores *válidos, fiables y apropiados*, como resume la Tabla 14.5.

TABLA 14.5. Características de un buen indicador

1. VALIDEZ

1.1. VALIDEZ COMO INDICADOR DE CALIDAD
¿Mide calidad y sirve para monitorizar e identificar situaciones en las que la calidad asistencial puede mejorarse?

1.2. VALIDEZ "FACIAL".
¿Se entiende su sentido e importancia sin muchas explicaciones?

1.3. SENSIBILIDAD.
¿Identifica todos los casos en los que hay problemas de calidad?

1.4. ESPECIFICIDAD
¿Identifica sólo los casos en los que hay un problema de calidad?

2. FIABILIDAD

¿Es interpretado siempre de la misma manera por todos los evaluadores?

3. UTILIDAD (PARA LA MEJORA DE LA CALIDAD)

¿Es apropiado para el nivel de responsabilidad de quienes valoran sus resultados?

6.1. INDICADORES VÁLIDOS

Como tal instrumento de *screening*, un buen indicador será aquél que reúna las características de validez, sensibilidad y especificidad. Por otra parte, un indicador es también un criterio de calidad, aunque sea un criterio especialmente relevante al tener vocación de resumen del nivel de calidad; por ejemplo, para el programa de hipertensión podríamos decidir que sería suficiente la selección de tres indicadores para monitorizar su calidad: uno en relación a la captación, otro en relación al diagnóstico correcto, y otro en relación al resultado. El que los indicadores sean también criterios de calidad quiere decir que todas las recomendaciones y características que vimos en la UT 8 para los criterios (realistas, aceptables, medibles, fiables; además de relevantes y válidos), son igualmente aplicables a los indicadores.

6.2. INDICADORES FIABLES

Un buen indicador ha de tener las mismas características exigibles a un buen criterio, más las que corresponde a su condición como instrumento de screening (sensibilidad y especificidad)

Para todos los criterios de calidad en general, si la validez asegura que la herramienta mida lo que queremos medir, la fiabilidad se define como el grado de reproducibilidad de los resultados para los mismos casos y situaciones cuando el indicador es utilizado por observadores diferentes. Es decir, que cada uno de los evaluadores obtenga el mismo resultado que el otro o los otros evaluadores, al evaluar la misma cosa con el mismo indicador. Sin fiabilidad no hay validez, porque estamos asignando valores a la medición de una forma inconsistente.

La realización de un pilotaje anterior a la generalización del uso del indicador, es considerada un paso inexcusable para asegurar la fiabilidad. El pilotaje permite medir la fiabilidad (concordancia, índice kappa, acuerdo específico) e identificar, discutir y corregir, si las hubiere, diferencias en la interpretación de los indicadores. El proceso de pilotaje debe repetirse tantas veces como sea necesario, hasta alcanzar cotas de fiabilidad aceptables. Una vez asegurada la fiabilidad, puede valorarse la sensibilidad y especificidad del indicador.

Para asegurarnos que construimos o seleccionamos un indicador que tenga las características de *validez* y *fiabilidad*, la Joint Commission americana propuso hace algunos años un esquema resumible en 7 puntos que hay que considerar explícitamente uno a uno, a modo de lista de comprobación o examen del indicador en cuestión. Ente listado figura en la Tabla 14.6, y su significado, de forma resumida, es el siguiente:

riación del indicador, de forma que sepamos qué tipo de problema va a detectar y qué tenemos que evaluar cuando la medición nos indique calidad defectuosa.

- **Comprobación empírica de la validez.** No siempre es posible, pero sí deseable fundamentar empíricamente la reflexión sobre los factores subyacentes que pensamos mide el indicador.

Una explicación más detallada de este esquema puede verse en el artículo de la Joint Commission en el que se basa, publicado en español por la revista de la Sociedad Española de Calidad Asistencial, y que figura en la bibliografía de esta UT. La aplicación del esquema de la Joint Commission ayuda a seleccionar buenos indicadores, que habrá que someter adicionalmente a pruebas de *fiabilidad* (como vimos en la UT 8). Sin embargo, un esquema más reciente desarrollado en *el Veterans Affairs Center for Practice Management and Outcomes Research*, también en Estados Unidos, contempla un proceso de validación de cuatro pasos y con más énfasis en la necesidad de pilotaje y comprobación empírica de diversos aspectos de la validez y fiabilidad, tal como se resume en la Tabla 14.7. En este esquema se contempla también de alguna forma (en el paso 4) la tercera y muy importante característica que ha de tener un indicador para incluirlo en un plan de monitorización: que sea *apropiado*.

Para analizar y construir un indicador es conveniente utilizar un listado estructurado de aspectos a definir o comprobar, como el propuesto por la Joint Commission

TABLA 14.7. Proceso a seguir para la construcción de buenos indicadores.

1. Selección de indicadores potenciales
   - 1.1. Valorar la evidencia científica que justifica la importancia del indicador.
   - 1.2. Valorar la facilidad de medición y la frecuencia de casos problemáticos que identifica el indicador.
   - 1.3. Valorar y clasificar los factores modificables que se asocian con los casos problemáticos que identifica el indicador.
2. Establecer estándares y diseñar herramientas de medición
3. Pilotaje

   Medición del indicador en grupos de casos y controles para valorar sensibilidad, especificidad y valor predictivo. Valoración de la fiabilidad.
4. Simulación de la aplicación del indicador en instituciones y niveles para los que se ha diseñado

   Valorar su aplicabilidad y utilidad real.

Adaptado de: Hofer TP., Bernstein SJ., Hayward Ra, De Monner S. Validating Quality Indicators for Hospital Care. Jt. Comm J. Qual Improv. 1997; 23 (9): 455-467.

6.3. INDICADORES APROPIADOS

En los programas de gestión de calidad no tiene sentido la monitorización como actividad aislada, es preciso reaccionar a sus resultados; por eso, además de válido y fiable, el indicador debe ser útil para la gestión de la calidad en la institución o nivel dentro del sistema de salud en el que vaya a ser utilizado; es decir, tiene que ser *apropiado* para que de los resultados de la monitorización puedan derivarse, en su caso, acciones de mejora. Un indicador no es apropiado para un determinado programa de gestión de calidad de un determinado

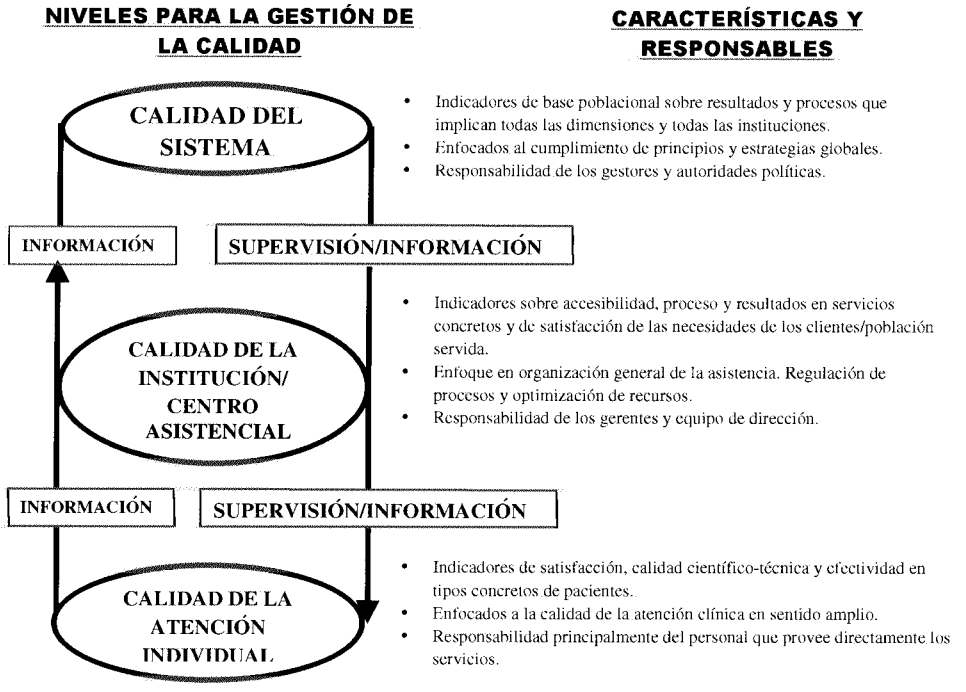
nivel dentro del sistema de salud si los resultados de su monitorización no pueden seguirse de acciones para la mejora *por parte de quienes lo están utilizando* cuando un problema es detectado

En la industria, las ideas más sugerentes sobre este tema son probablemente las que aporta Juran cuando explica el concepto de la "pirámide de control", como representación de las cosas que una determinada empresa debe monitorizar en relación a sus productos y procesos en aras a mantener y mejorar su calidad; es una especie de "plan de delegaciones" cuya base son los controles automatizados, y los sucesivos escalones, cada vez más estrechos, son los controles que ha de ejercer la mano de obra, los supervisores y mandos intermedios, y, finalmente, la cúspide del nivel gestor; la idea es que el papel de los altos directivos ha de ser exclusivamente el control sobre los grandes objetivos estratégicos y asegurarse de que existe y funciona un sistema de control en el resto de los segmentos de la pirámide. Cada nivel debería controlar indicadores relacionados con su responsabilidad en el sistema.

Validez, fiabilidad y utilidad son las tres características a exigir a los indicadores para monitorizar la calidad.

Sobre la base de estas ideas, adaptándolas a lo que serían las características de un indicador apropiado, según el nivel/responsabilidad dentro del sistema de salud en el que se quiera utilizar, proponemos el esquema que se detalla en la Figura 14.2, como punto de partida para un tema que se sabe importante, pero que está aun sin resolver de forma satisfactoria en prácticamente ningún sistema de salud. Dos ejemplos reales de intento de poner en práctica esta clara delimitación de responsabilidades y niveles para la monitorización son la forma en que están organizados los programas de Gestión de la Calidad en la red de establecimientos (en la práctica un sistema de salud) para los que han trabajado en el ejército de Estados Unidos (Veteran Health Administration), con tres niveles establecidos (nacional, regional y de centro hospitalario) y la reciente iniciativa del Servicio Nacional de Salud Británico que distingue y define indicadores a nivel de Región y centros hospitalarios.

FIGURA 14.2. Calidad e indicadores en los distintos niveles del sistema



7. COMPONENTES DE UN PLAN DE MONITORIZACIÓN. TIPOS DE PLAN

Una vez definidos y seleccionados los indicadores, sea cual sea la situación o punto de partida (Tabla 14.3), las principales características a describir en el plan de monitorización son: (i) la frecuencia de las mediciones; (ii) los mecanismos de recogida de datos; y (iii) los métodos que guían el diseño de esta recogida y la interpretación de los mismos.

La decisión sobre la *frecuencia* de la medición puede determinar en gran medida otras características del plan de monitorización, incluido el método o métodos que pueden ser de mayor utilidad. Cada situación y tipo de indicador puede requerir mediciones con mayor o menor frecuencia. En los servicios de salud, lo más común en los programas externos existentes en la actualidad es realizar mediciones anuales o, cuando menos, mensuales pero puede darse la necesidad u oportunidad de realizar mediciones más frecuentes. En relación a la frecuencia de las mediciones podemos distinguir planes con frecuencia elevada (*continuos* o de intervalos cortos entre mediciones), y planes *espaciados*, con mediciones a intervalos largos. Vamos a considerar, de una forma un tanto arbitraria, un mes como límite de frecuencia entre ambos tipos de planes.

Los planes de monitorización incluyen la frecuencia de las mediciones, o mecanismos de la recogida de datos, y los métodos para su interpretación. De todo ello, lo que más va a influir en la práctica concreta es la frecuencia de mediciones y el interés que podamos tener en estimar el nivel de calidad.

Una característica común a cualquier plan de medición que se realice con muestras es la necesidad de que éstas sean aleatorias, debiéndose decidir y definir, como en las evaluaciones, el *tipo de muestreo* (aleatorio simple, sistemático, estratificado, etc.) y la *mecánica de obtención* de las unidades de estudio en las que mediremos el indicador. Mecánica que ha de estar en relación con el *método* de monitorización elegido.

En relación a esto último, tal como se resume en la Tabla 14.2, siempre está la opción de enfrentarse a las mediciones del indicador como si fuesen estimaciones del nivel de cumplimiento del mismo cada vez que lo midamos, y comparar nuestros resultados con el estándar o nivel de cumplimiento de referencia; la metodología y análisis de las mediciones sería entonces semejante a la que se explica en las UT 11 (estimación del nivel de calidad) y UT 13 (comparación de dos evaluaciones). Sin embargo, si renunciamos a conocer el nivel real de cumplimiento del indicador que medimos, y centramos nuestra atención en saber si la situación es o no problemática, podemos utilizar métodos desarrollados en la industria, que tienen la ventaja de utilizar tamaños de muestra más pequeños, además de poder servir, en el caso de los gráficos de control estadístico, para un conocimiento más profundo del tipo de problema que pueda existir.

8. MÉTODOS DE MONITORIZACIÓN SEGÚN LAS CARACTERÍSTICAS DEL PLAN DE MONITORIZACIÓN

Centrándonos en los métodos desarrollados en la industria, podemos distinguir, siguiendo su propia terminología, dos categorías: métodos para *inspección de la calidad*, y métodos para el *control de procesos*. La inspección se define como la medición de unos determinados indicadores y su comparación con los requisitos o estándares de calidad para determinar su cumplimiento o su conformidad con ellos. El control de procesos implica la medición y análisis de la variación de los indicadores seleccionados, que han de representar aspectos relevan-

tes del proceso sometido a análisis y control. Los planes de medición implican, en ambos casos, mediciones sistemáticas y repetidas con una periodicidad dada que es normalmente mayor para el control de procesos.

La inspección es contemplada por algunos como un método transversal y estático orientado a medir la calidad en una determinada población (grupo o "lote") del cual se extrae la muestra, para decidir si aceptar o no que su calidad es adecuada; mientras que el control de procesos es más una actividad continua y dinámica para tomar decisiones sobre el proceso controlado. Estas decisiones son, esencialmente, en el sentido de aceptar o rechazar si hay que intervenir, o investigar una variabilidad que puede resultar atípica o inesperada, lo que le asemeja al tipo de decisión (aceptación/rechazo de situación problemática) que se realiza en la inspección; sin embargo, existe hoy día una discusión abierta sobre la conveniencia de realizar o no "inspecciones" dado que el énfasis debe ponerse en el control de procesos y los programas de prevención que consigan los niveles de "conformidad" deseados. Esta discusión no es, sin embargo, sobre los métodos en sí, sino sobre la oportunidad y objetivos al aplicar uno u otro enfoque.

El muestreo para la aceptación de lotes (*Lot Quality Acceptance Sampling*, LQAS) y los gráficos de control estadístico de la calidad representan los métodos de *inspección* y de *control de procesos* que pueden ser utilizados de forma eficiente en la monitorización de la calidad en los servicios de salud. A pesar de las diferencias conceptuales que podamos encontrarles, ambos tienen en común dos características relevantes:

- **No proporcionan como resultado estimaciones del nivel de calidad**. En su lugar informan, de maneras diferentes, sobre diferencias estadísticamente significativas en relación al nivel o condiciones predeterminadas. Esta información se utiliza para clasificar como aceptable o problemático al proceso, producto o servicio bajo escrutinio.

  El que no sean útiles para dar estimaciones del nivel de calidad puede ser visto como una desventaja para el personal de los servicios de salud, acostumbrados a trabajar con estimaciones de niveles, prevalencias, cifras que miden los diversos parámetros fisiológicos, etc. Sin embargo, estos métodos nos dicen que tener estimaciones del nivel de calidad no es necesario para cumplir con el objetivo de la monitorización, si la consideramos como una actividad para el screening o identificación de problemas. La renuncia a tener estimaciones es una de las condiciones para la segunda característica de los métodos de control de la calidad en la industria.

- **Tamaño de muestra.** Permiten tomar decisiones utilizando *tamaños de muestra* mucho más *pequeños* que los que necesitaríamos si quisiésemos decidir por medio de comparar una estimación del nivel de calidad al estándar prefijado, proceder este último que constituye, por otra parte, la forma de actuar más común e intuitiva en la tradición de los servicios de salud. Las decisiones sobre la aceptación (tras *inspecciones*) o sobre la existencia de variación problemática (a través del *control de procesos*) pueden tomarse con tamaños de muestra relativamente pequeños, por medio de una utilización especialmente práctica e inteligente de la estadística, y más en concreto de la teoría de la probabilidad.

Si renunciamos a estimar el nivel de calidad, podemos utilizar métodos desarrollados en la industria como el LQAS y el control gráfico de la calidad, que utilizan en general muestras más pequeñas.

Sin embargo, al margen de las discusiones teóricas, hay una diferencia importante en la aplicación *práctica* de ambos métodos: el control gráfico sólo es

aplicable en planes de medición continua o de intervalos cortos entre mediciones, mientras que el LQAS puede utilizarse en cualquier tipo de plan, sea con mediciones a intervalos cortos o largos, entendiendo por intervalo corto, tal como apuntábamos más arriba, si es mensual o menor. La Tabla 14.8 resume esta distinción, sobre la que queremos subrayar que los métodos de control gráfico son tanto más útiles cuánto más frecuentes (menos espaciadas) sean las mediciones, y que el LQAS es probablemente el método de elección para monitorizar y priorizar problemas de calidad si esta actividad se realiza cada varios meses, anualmente o de forma esporádica. El conocimiento detallado de la práctica de ambos métodos hace evidentes las razones de esta distinción, como veremos en las UT siguientes.

TABLA 14.8. Planes y métodos de monitorización según frecuencia de las mediciones

PLAN (según frecuencia de mediciones)	MÉTODOS
• ESPACIADO (intervalos largos* o esporádicos)	• EVALUACIÓN RÁPIDA • LQAS**
• CONTINUO (intervalos cortos)	• GRÁFICOS DE CONTROL ESTADÍSTICO DE LA CALIDAD • LQAS

*: Intervalo largo: Mediciones con frecuencia superior a un mes
**: LQAS: Lot Quality Acceptance Sampling= Muestreo para la aceptación de la calidad de lotes.

BIBLIOGRAFÍA

- Davins J. Construcción y analisis de indicadores para monotorizar la Calidad. En: Saturno PJ, Gascón JJ, Parra P. Calidad Asistencial en Atención Primaria. Tomo II. Madrid: Dupont Pharma; 1997. p. 249-268.
- Saturno PJ. Planes de monitorización. Muestreo para la aceptación de lotes. En: Saturno PJ, Gascón JJ, Parra P. Calidad Asistencial en Atención Primaria. Tomo II. Madrid: Dupont Pharma; 1997. p. 269-344.
- Saturno PJ. Control estadístico de la Calidad. Monitorización con gráficos de control. En: Saturno PJ, Gascón JJ, Parra P. Calidad Asistencial en Atención Primaria. Tomo II. Madrid: Dupont Pharma; 1997. p. 305-344.
- Joint Commission on Acreditation of Health Care Organization (JCAHO). Características de los indicadores clínicos. Control de Calidad Asistencial 1991; 6: 65-79.
- Saturno P.J. Qué, cómo y cuando monitorizar: Marco conceptual y guía metodológica. Revista de Calidad Asistencial 1998; 13: 437-443.
- Halpern J. The measurement of quality of care in the Veterans Health Administration. Med Care 1996; 34 (3): M555-M568.
- Quality and Performance in the NHS; High Level Performance Indicators and Clinical Indicators. London: NHS Executive; 1999. Disponible en: URL:http://www.doh.gov.uk/indicat/

EL MUESTREO PARA LA ACEPTACIÓN DE LOTES (LQAS) COMO MÉTODO DE MONITORIZACIÓN

EMCA
Gestión de la Calidad Asistencial

CONTENIDO GENERAL

Después de aclarar el concepto, situaciones en que es útil y los objetivos de la monitorización, así como la importancia y la forma de seleccionar buenos indicadores, en esta UT se abordan las bases teóricas y la aplicación práctica del LQAS. Comenzamos por revisar los fundamentos y aplicación de la distribución binomial en la evaluación rápida de la calidad con muestras pequeñas, para después, sobre esta base, entender y practicar el LQAS en planes de monitorización con intervalo largo entre mediciones.

ÍNDICE DE CONTENIDOS

1. Introducción.
2. Planes de mediciones esporádicas o con grandes intervalos de tiempo: la "evaluación rápida" de la calidad.
3. El muestreo de aceptación de lotes como método de monitorización en los servicios de salud.
4. Índices de calidad, riesgos y procedimientos para la aplicación del LQAS.
5. LQAS en servicios de salud: dos ejemplos prácticos.
6. Esquemas avanzados de muestreo de aceptación de lotes.

OBJETIVOS ESPECÍFICOS

1. Entender el fundamento estadístico de la aplicación de la distribución binomial a la evaluación rápida de problemas de calidad.
2. Interpretar en términos de probabilidad estadística el resultado de una muestra pequeña (<30 casos) sobre un criterio o indicador dicotómico.
3. Distinguir en la práctica los conceptos de errores α y β en el contraste de hipótesis.
4. Entender el fundamento estadístico del LQAS como método de medición.
5. Diseñar planes de monitorización que incluyen el LQAS como método de medición.
6. Manejar las tablas LQAS, de acuerdo con unas condiciones de referencia (estándar, umbral, errores α y β) prefijadas.
7. Conocer diversos esquemas y variaciones del LQAS, según coste, métodos de muestreo y tipo de probabilidad aplicados.

1. INTRODUCCIÓN

Según el concepto que vimos en la UT 14, entendemos por monitorización la medición sistemática, repetida y planificada de indicadores de calidad; una actividad conducente a controlar que estamos a unos niveles preestablecidos y que tiene como objetivo identificar la existencia o no de situaciones problemáticas que hay que evaluar o sobre las que hay que intervenir. Los indicadores a monitorizar pueden derivarse de ciclos de mejora, de actividades de diseño de los servicios, o ser fruto de una selección de aspectos o servicios relevantes de nuestro centro cuya calidad nos interesa controlar. En cualquier caso, deben ir acompañados de un plan de monitorización que incluya periodicidad, mecanismos para la recogida de datos y método de interpretación de los mismos.

El interés que podamos tener en una estimación del nivel de calidad existente y la periodicidad de medición son decisiones de importancia para la elección de los métodos para monitorizar. El muestreo de aceptación de lotes (LQAS) y el control estadístico (gráfico) de la calidad son métodos originados y desarrollados en la industria, que pueden resultar de gran utilidad para la monitorización en servicios de salud si renunciamos a tener una estimación del nivel de cumplimiento de los indicadores y nos centramos en el conocimiento de la existencia o no (aceptación/rechazo) de unos niveles de cumplimiento preestablecidos. El control estadístico de la calidad precisa de mediciones frecuentes, (mínimo mensuales) para ser útil, mientras que el LQAS y otros métodos de evaluación rápida basados en la distribución binomial pueden ser aplicados a planes de mediciones más espaciados. La práctica de monitorización utilizando LQAS, aún poco extendida en los servicios de salud, es, sin embargo, muy sencilla; sólo precisa entender y manejar unas tablas ya construidas, para fijar cuál es el número máximo de incumplimientos del indicador que es aceptable en una muestra determinada (y normalmente pequeña) para aceptar a su vez un determinado nivel de cumplimiento preestablecido, con unos niveles de error estadístico previamente acordados. Esta UT ofrece los fundamentos estadísticos y el detalle de las implicaciones prácticas del uso del LQAS en servicios de salud.

En esta UT se ofrecen los fundamentos teóricos y la aplicación práctica de la distribución binomial y el LQAS, para la evaluación rápida de problemas y la monitorización de indicadores en servicios de salud

2. PLANES DE MEDICIONES ESPORÁDICAS O CON GRANDES INTERVALOS DE TIEMPO: LA "EVALUACIÓN RÁPIDA" DE LA CALIDAD

Cuando la variable que mide el indicador seleccionado es de tipo dicotómico (presencia/ausencia de una determinada cualidad), como ocurre en la mayoría de los indicadores de calidad que se construyen en los servicios de salud, la utilización de las probabilidades de la distribución binomial nos proporciona una forma rápida y eficiente de decidir si estamos o no ante una situación problemática, utilizando un tamaño de muestra relativamente pequeño (máximo 30 casos). Vamos a ver cómo funciona y cuáles son las circunstancias en que es aplicable, partiendo de dos supuestos prácticos.

Supongamos que, refiriéndonos a dos problemas de salud diferentes, nuestro estándar histórico de cumplimiento de toma de tensión arterial para la detección de problemas de hipertensión se sitúa en el 85%, y que la prescripción de antibióticos en casos de resfriado común hemos logrado reducirla al 15 % de los

casos. Sin embargo, en la actualidad, después de habernos despreocupado durante varios años, se percibe estancamiento en el número de pacientes clasificados como hipertensos, lo cual nos hace sospechar que no estamos haciendo a buen nivel las actividades de detección; y por otra parte, hace tiempo que no comprobamos si mantenemos el nivel de prescripción inadecuada de antibióticos a ese bajo nivel al que tanto nos costó llegar. En ambos casos nos interesa saber si la situación es o no problemática para decidir si debemos o no evaluar qué está pasando o intervenir sobre ella. Una opción sería realizar un muestreo aleatorio suficientemente grande de las visitas en los últimos 6 meses tanto de pacientes en general (para evaluar la detección de hipertensos), como, de forma específica, de los que fueron diagnosticados de resfriado común (para evaluar la prescripción de antibióticos), y estimar el nivel al que nos encontramos.

Las tablas de distribución binomial pueden utilizarse para decidir de una forma rápida y con una muestra pequeña si existe o no problema de calidad en relación a un estándar preestablecido y siempre que el indicador se mida con una variable dicotómica (de dos opciones: cumple/no cumple).

Sin embargo, ambas situaciones cumplen los requisitos en los que es aplicable la distribución binomial a saber:

- Lo que se evalúa tiene sólo dos valores posibles (cumple/no cumple)
- Cada una de las unidades de estudio es independiente de la otra.
- Evaluamos en una muestra de "n" casos.
- La probabilidad general que existe de encontrar un valor u otro (cumple/no cumple) es la misma para cada uno de los casos de la muestra.

El tamaño de la muestra ha de ser inferior al 10% del tamaño del marco muestral.

Esta última condición se cumple siempre que el tamaño de la muestra "n", sea inferior al 10% del universo o marco muestral de donde la hemos extraído (en nuestro caso el total de visitas en los últimos seis meses, y el total de casos de resfriado común respectivamente). Si "n" es mayor que el 10% del marco muestral habría que hacer un tipo de muestreo especial (el llamado muestreo con reposición) o utilizar la distribución hipergeométrica, de la cual no nos vamos a ocupar en esta UT.

En la que llamamos "evaluación rápida" interpretamos el resultado de la muestra en términos de la probabilidad de que este pueda darse si fuera cierto que se cumple el estándar prefijado en la población o marco muestral.

Si es aplicable la distribución binomial, vamos a extraer una muestra aleatoria pero pequeña, por ejemplo 15 casos, e interpretar el resultado que obtengamos en términos de la probabilidad de que sea compatible con la existencia real del estándar prefijado (85% de cumplimiento para la detección de hipertensión, 15% para la prescripción de antibióticos), sin que nos interese conocer cual es el nivel preciso de cumplimiento que existe; sólo queremos saber si podemos aceptar o no que estamos a los niveles que queremos tener.

Evaluamos nuestras dos muestras de 15 casos y encontramos que para la hipertensión el número de casos en que se incumple el indicador es de 5, y en 4 de los casos de resfriado común se han prescrito antibióticos. ¿Qué podemos deducir de estos resultados? Obviamente no es apropiado estimar el nivel de cumplimiento con una muestra tan pequeña, pero lo que si podemos hacer es averiguar cuál es la probabilidad de que esos resultados sean compatibles con la existencia del estándar prefijado en la población de la que hemos extraído la muestra. Si esta probabilidad es baja, concluiremos que este resultado es incompatible con el estándar prefijado y que por tanto rechazamos que este estándar se cumpla en la población de la que hemos extraído la muestra; según el consenso más extendido en las pruebas de significación estadística podemos considerar que una probabilidad es baja si es $\leq 5\%$.

Para hallar la probabilidad que nos interesa (en nuestro caso la que tiene el

encontrar 5 incumplimientos de toma de tensión en una muestra de 15 si el estándar es de 85% de cumplimiento, y el encontrar 4 casos de prescripción de antibióticos en una muestra de 15, si el estándar es de 15%), podemos utilizar la fórmula de probabilidades de la distribución binomial, (cuyo cálculo adelantamos que no es preciso efectuar porque existen las tablas oportunas):

$$P_{(x)} = \begin{bmatrix} n \\ x \end{bmatrix} \pi^{x} (1 - \pi)^{n-x}$$

en donde:

P (x): es la probabilidad de encontrar casos de cumplimiento (o incumplimiento, según hayamos definido el estándar) en la muestra.

n: es el tamaño de la muestra

x: son los casos de cumplimiento (o incumplimiento) que hemos encontrado en la muestra

p: es la probabilidad de cumplimiento (o incumplimiento) prefijada (estándar), que suponemos existe en la población de la que extraemos la muestra.

$\begin{bmatrix} n \\ x \end{bmatrix}$: es una expresión combinatoria que equivale a : $\frac{n!}{x!(n-x)!}$

siendo la anotación

"!" indicativa de una multiplicación factorial.

Es importante señalar que los valores de π y x han de ir *en el mismo sentido;* es decir, si el estándar es de cumplimiento (como en nuestro ejemplo de la hipertensión) x ha de ser número de cumplimientos, pero si el estándar es negativo o de incumplimiento, x deben ser incumplimientos. En nuestros dos ejemplos *n* es igual a 15 en ambos casos; π es igual a 85% para el caso de la hipertensión y 15% para la prescripción de antibióticos; x sería 10 para la hipertensión (hemos encontrado 5 incumplimientos, pero el estándar lo tenemos en positivo), y 4 para la prescripción de antibióticos en el resfriado común.

Para averiguar P(x) (la probabilidad de haber encontrado precisamente el número de cumplimientos que hemos encontrado si el estándar se cumple en la población de dónde hemos extraído la muestra) podríamos utilizar la fórmula descrita que, aunque parezca complicada, es fácil de programar en cualquier hoja de cálculo; sin embargo es más cómodo utilizar las tablas de la distribución binomial existentes en los libros de estadística. Hay una tabla de probabilidades para cada tamaño de muestra (normalmente hasta n = 20 ó, como máximo n =30).

Estas tablas dan los cálculos de probabilidad que nos interesan para cada uno de los posibles resultados en un determinado tamaño de muestra. En consecuencia, lo que tenemos que hacer en primer lugar es localizar la tabla que *corresponde al tamaño de muestra que hemos utilizado.*

En nuestro caso ha sido 15, y la tabla correspondiente a n =15 es la que reproducimos como Tabla 15.1. En ella podemos ver que las columnas están encabezadas por unas cifras que corresponden a los diversos valores (de .05 a .50) que podemos hacer corresponder con el estándar prefijado. Para localizar las probabilidades que nos interesan en nuestros ejemplos nos situaremos en la columna de π (estándar) igual a 0,15 (15%), tanto para el caso de la prescrip-

La probabilidad del resultado encontrado se busca en las tablas correspondientes de la distribución binomial para el tamaño de muestra empleado. Normalmente se rechaza que sea cierto el cumplimiento del estándar de referencia si la probabilidad del resultado obtenido es ≤ 0,05.

Para buscar en las tablas hay que localizar primero la correspondiente al tamaño de la muestra, y después mirar en la columna del estándar correspondiente

ción de antibióticos (que ya está expresado así : el indicador busca la no prescripción de antibióticos), como en el caso de la hipertensión (un estándar de cumplimiento de 85% es equivalente a un estándar de incumplimiento de 15%).

A continuación, buscamos en la columna de "x" el número de incumplimientos que hemos encontrado: 5 para la hipertensión y 4 para la prescripción de antibióticos. El cruce de la fila de este número con la columna de 0,15 expresa la probabilidad de encontrar precisamente ese número de incumplimientos si el estándar se cumple. En la Tabla 15.1 podemos ver que estas probabilidades son 0,0449 (4,49%), para el caso de la hipertensión, y 0,1156 (11,56%) para el caso de los antibióticos. ¿Cómo interpretarlo?: ateniéndonos a la norma para decidir establecida (rechazar que sean ciertas las situaciones de cumplimiento del estándar si el resultado de la muestra tiene una probabilidad ≤ 5%), identificamos como problemática la situación de la detección de hipertensos, mientras que aceptamos como no problemática (en relación a nuestro estándar) la prescripción de antibióticos en el resfriado común.

TABLA 15.1. Tabla de probabilidades binomiales (para n=15)

	ESTÁNDAR DE INCUMPLIMIENTOS (π)									
x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
1	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
2	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
3	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
4	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
5	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
6	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
7	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
8	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
9	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
10	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
11	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
12	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0139
13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

x= nº de incumplimientos encontrados en la muestra de 15 casos.

Fte: Adaptada de Rosner

Este es un proceder sencillo, que sólo precisa de:

1. Establecimiento de unos parámetros de referencia:
   - Estándar que queremos comprobar si se cumple o no (π), para el indicador a comprobar.

- Un tamaño de muestra *(n)* que nos resulte manejable para tomar una decisión rápida.

2. Extracción de la muestra de forma aleatoria.

3. Buscar en las tablas de la distribución binomial la probabilidad de encontrar el resultado que hayamos obtenido en la muestra, en la columna correspondiente al valor del estándar cuyo cumplimiento queremos comprobar. Si esta probabilidad es baja, clasificamos la situación como problemática y la investigamos más a fondo.

La distribución binomial no tiene generalmente utilidad práctica para valorar muestras de más de 30 casos.

Es muy importante asegurarse de entender bien los diversos parámetros que utilizamos, porque las tablas de la distribución binomial que contienen los diversos libros de estadística no utilizan siempre los mismos símbolos. Por ejemplo, "x" puede aparecer como "a", o como "c", o como "k". Por otra parte, las probabilidades a veces son exactas (como en la Tabla 15.1) o acumuladas; π puede estar en positivo (cumplimientos) o negativo (incumplimientos). La estructura de la tabla es, sin embargo, siempre la misma.

Este proceder de "evaluación rápida" es muy sencillo y práctico para interpretar muestras pequeñas en términos de decidir la aceptación o rechazo sobre la existencia de un determinado estándar. Adviértase que el rechazo puede indicar identificación de situación problemática o también que la situación sea mejor que el estándar. Véase, por ejemplo, en la Tabla 15.1 que para un estándar de 30%, es prácticamente igual de improbable encontrar 1 incumplimiento que 8 incumplimientos en una muestra de 15 casos, y con ambos resultados rechazaremos que exista estandar del 30%; sin embargo, en el primer caso (1 incumplimiento) lo más probable es que en realidad la tasa de incumplimiento sea menor, y en el segundo caso (8 incumplimientos) que sea mayor que el estándar de 30% de incumplimiento que utilizamos como referencia.

La que hemos llamado "evaluación rápida" sólo puede utilizarse para monitorizaciones esporádicas (no sistemáticas) y sólo tiene en cuenta uno de los dos posibles errores que pueden darse en el contraste de hipótesis.

Sin embargo, a pesar de su sencillez, la "evaluación rápida" tiene una serie de *limitaciones* y utilidades que es preciso conocer, para no utilizarla de forma inadecuada. Las limitaciones (Tabla 15.2) son su inadecuación para monitorizaciones sistemáticas y su parcialidad como método de contraste de hipótesis, que veremos a continuación.

TABLA 15.2. Utilidad y limitaciones de la "evaluación rápida" binomial como método de identificación de problemas con indicadores tipo tasa (proporción o porcentaje)

UTILIDAD	LIMITACIONES
Monitorización esporádica (no sistemática) de indicadores.	Inadecuada para monitorización sistemática, aunque puede ser adaptada para ello.
Interpretación de resultados de muestras pequeñas (<30 casos), en relación a un estándar de referencia.	Es incompleta como método de contraste de hipótesis: la decisión solo tiene en cuenta la probabilidad puntual de falsos positivos (identificar como problemática una situación que no lo es) o de falsos negativos (aceptar como buena una situación problemática), pero no ambos.

2.1. MONITORIZACIÓN SISTEMÁTICA UTILIZANDO LA DISTRIBUCIÓN BINOMIAL

Para monitorizar de forma sistemática utilizando la distribución binomial deben utilizarse las tablas de probabilidad acumulada.

Los resultados de una evaluación puntual, esporádica, pueden ser interpretados en la forma que lo hemos hecho con las tablas de distribución binomial, pero no sería correcto mantener el mismo proceder como método de evaluación sistemática, repetida o rutinaria. En este caso lo correcto sería considerar para nuestra decisión, no la probabilidad del resultado concreto que hemos encontrado en la muestra puntual (tal como hemos hecho) sino la probabilidad de todos los resultados posibles *simultáneamente* (probabilidad acumulada) para el tamaño de muestra que hemos decidido utilizar. Es decir, establecemos de igual manera una probabilidad límite para aceptar o rechazar (que puede seguir siendo $\leq 0,05$), pero su cálculo implica la suma de las probabilidades de que encontremos un determinado número de incumplimientos y cualquier número de incumplimientos inferior. La decisión no se tomaría entonces en función de la probabilidad para un resultado concreto (como hemos hecho en nuestros ejemplos), sino que buscaríamos un número *límite* de incumplimientos de forma que la probabilidad de encontrar ese número de incumplimientos o más sea $\leq 0,05$. Sólo aceptaríamos como buenos los resultados con número de incumplimiento por debajo de ese límite. Si en vez de incumplimientos quisiéramos valorar cumplimientos el razonamiento es semejante aunque naturalmente, en sentido opuesto: buscaríamos un número mínimo de cumplimientos de forma que la probabilidad de encontrar ese número o más fuese $\geq 0,95$; así identificaríamos como problemático todo resultado que tuviera un número de cumplimientos inferior al que hemos señalado como límite para la decisión. La decisión se toma, como se ve, en función de la probabilidad *acumulada* de los diversos resultados posibles. Para ello debemos utilizar las tablas de la distribución binomial con probabilidades acumuladas, tal como la que se reproduce también para una muestra de 15 casos, en la Tabla 15.3.

TABLA 15.3. Tabla de probabilidades binomiales acumuladas (para n=15)

	ESTÁNDAR DE INCUMPLIMIENTOS (π)									
x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
1	.8290	.5490	.3186	.1671	.0802	.0353	.0142	.0052	.0017	.0005
2	.9638	.8159	.6042	.3980	.2361	.1268	.0617	.0271	.0107	.0037
3	.9945	.9444	.8227	.6482	.4613	.2969	.1727	.0905	.0424	.0176
4	.9994	.9873	.9383	.8358	.6865	.5155	.3519	.2173	.1204	.0592
5	.9999	.9978	.9832	.9389	.8516	.7216	.5643	.4032	.2608	.1509
6	1.0000	.9997	.9964	.9819	.9434	.8689	.7548	.6098	.4522	.3036
7	1.0000	1.0000	.9994	.9959	.9827	.9500	.8868	.7869	.6535	.5000
8	1.0000	1.0000	.9999	.9992	.9958	.9848	.9578	.9050	.8182	.6064
9	1.0000	1.0000	1.0000	.9999	.9992	.9963	.9876	.9662	.9231	.8491
10	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9972	.9907	.9745	.9408
11	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9981	.9937	.9824
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9989	.9963
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000.	1.0000

En la Tabla 15.3 las cifras de probabilidades que aparecen son la suma de la probabilidad de cada una de las posibles x, y de todas las anteriores. Por ejemplo, si nos colocamos en la columna que corresponde al estándar de incumplimiento de 15% (π=0.15) la probabilidad que encontramos para x=4 es 0.9383 (ó 93,83%) y corresponde a la suma de la probabilidad de que obtengamos, 0, 1, 2, 3, ó 4 incumplimientos en la muestra; es decir 4 incumplimientos o menos, mientras que en la Tabla 15.1 lo que veíamos es la probabilidad de obtener *exactamente* 4 incumplimientos.

Para utilizar el procedimiento de evaluación rápida en la monitorización sistemática, utilizaríamos las tablas de probabilidad acumulada, como la Tabla 15.3, buscando en la tabla correspondiente al tamaño de muestra que hayamos decidido utilizar, el número de incumplimientos que represente una probabilidad acumulada $\geq$ 0,95 y sea lo más cercano posible a este valor. Este número sería el que utilizaríamos para decidir, sabiendo que tendríamos una posibilidad de equivocarnos al clasificar como problemática la situación $\leq$ 0,05 (probabilidad de que encontremos un número de incumplimientos mayor del que utilizamos para clasificar la situación como problemática).

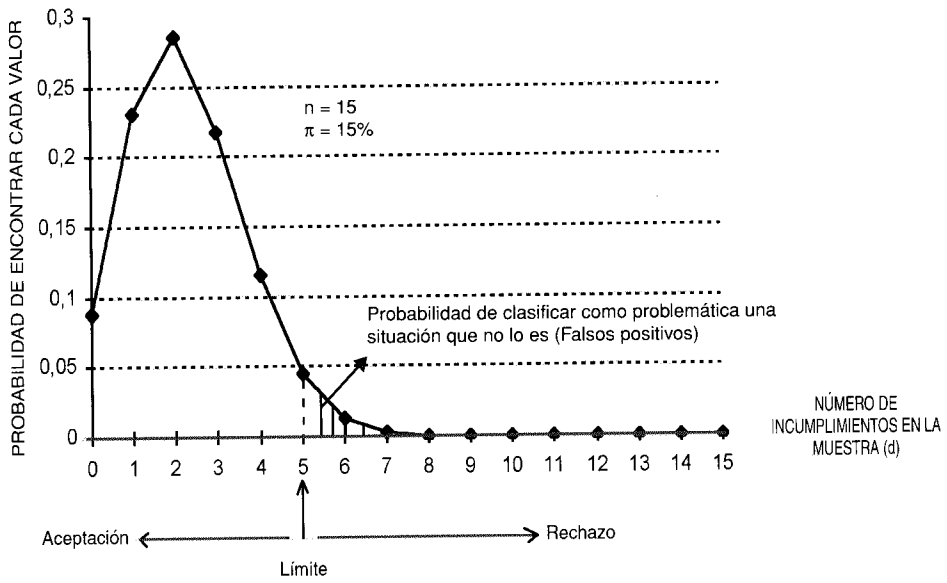
Para el caso de la prescripción de antibióticos (π =0,15), y con una muestra de 15 casos (Tabla 15.3), el número límite para decidir sería 5 (probabilidad acumulada = 0,9832). Clasificaremos como aceptables los resultados de 5 incumplimientos o menos, e inaceptables (problemáticos) si encontramos 6 incumplimientos o más; con ello la probabilidad de falsos positivos (decir que hay problema cuando no lo hay) que estamos aceptando es de 0,0168 (1-0,9832) que es la probabilidad de obtener más si el estándar es de 15%. Si fijamos "4"

como número de decisión, la tasas de falsos positivos que arriesgamos con nuestro proceder sería 1-0,9383=0,062 ó 6,2%. (0,9382 es la probabilidad de encontrar 4 incumplimientos o menos en una muestra de 15 casos, si el incumplimiento real es 15%).

La decisión para aceptar/ rechazar que el estándar se cumpla en la población o universo de donde se ha extraído la muestra, se base en que la probabilidad acumulada de obtener el resultado sea ≤ 0,05.

En la Figura 15.1 se ha representado la curva de probabilidades que contiene la Tabla 15.1 para el estándar de 15%. En esta figura puede observarse la probabilidad de todos los valores de incumplimiento posibles en una muestra de 15 casos, y las consecuencias en cuanto a falsos positivos (estándar real de 15% pero que clasificamos como problemático al situar en 5 el límite aceptable de incumplimientos). Aunque la representación como curva no es la más adecuada para las probabilidades binomiales, al no ser realmente continuas sino "a saltos" por no haber probabilidad de valores intermedios entre un número de incumplimientos y el siguiente, resulta más fácil de entender gráficamente y por eso la utilizamos.

FIGURA 15.1. Curva de probabilidades de los diversos valores de incumplimiento que podemos encontrar en una muestra de 15 casos extraida de una población con un nivel de incumplimiento = 15%



La Tabla 15.4 es semejante a la 15.3 (probabilidades acumuladas), pero π y x son cumplimientos, en vez de incumplimientos.

TABLA 15.4. Tabla de probabilidades binomiales acumuladas (para n=15)

	ESTÁNDAR DE CUMPLIMIENTOS (π)									
x	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0037	.0011	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000
3	.0176	.0063	.0019	.0005	.0001	.0000	.0000	.0000	.0000	.0000
4	.0592	.0255	.0093	.0028	.0007	.0001	.0000	.0000	.0000	.0000
5	.1509	.0769	.0338	.0124	.0037	.0008	.0001	.0000	.0000	.0000
6	.3036	.1818	.0950	.0422	.0152	.0042	.0008	.0001	.0000	.0000
7	.5000	.3465	.2131	.1132	.0500	.0173	.0042	.0006	.0000	.0000
8	.6964	.5478	.3902	.2452	.1311	.0566	.0181	.0036	.0003	.0000
9	.8491	.7392	.5968	.4357	.2784	.1484	.0611	.0168	.0022	.0001
10	.9408	.8796	.7827	.6481	.4845	.3135	.1642	.0617	.0127	.0006
11	.9824	.9576	.9095	.8273	.7031	.5387	.3518	.1773	.0556	.0055
12	.9963	.9893	.9729	.9383	.8732	.7639	.6020	.3958	.1841	.0362.
13	.9995	.9983	.9948	.9858	.9647	.9198	.8329	.6814	.4510	.1710
14	1.0000	.9999	.9995	.9984	.9953	.9866	.9648	.9126	.7941	.5367
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

En el ejemplo del indicador sobre captación de hipertensos, para el que hemos fijado un estándar de cumplimiento de 85% (π=0,85), y para su monitorización sistemática con una muestra de 15 casos buscamos en la Tabla 15.4 el número límite de cumplimientos cuya probabilidad acumulada sea $\leq$ 5%, es decir, que la suma de las probabilidades de todos los valores de cumplimientos por debajo del número límite, considerados en su conjunto sea $\leq$ 0,05. Mirando en la Tabla 15.4, ese número límite para aceptar o rechazar que el estándar se cumple es 10 (probabilidad acumulada de encontrar 9 ó menos = 0,0168). De manera que identificamos como problema siempre que el número de cumplimientos sea inferior a 10. Obsérvese que esta norma para monitorizar es la misma que hemos determinado para el caso de la prescripción de antibióticos, sólo que viendo cumplimientos, en vez de incumplimientos. No es de extrañar, porque un estándar de cumplimientos (en "positivo") de 85% es equivalente a uno de incumplimientos (en "negativo") de 15%, sólo que en un caso decidimos en base a un número mínimo de cumplimientos, y en el otro en base a un número máximo de incumplimientos.

2.2. CONTRASTE DE HIPÓTESIS "COMPLETO" UTILIZANDO LA DISTRIBUCIÓN BINOMIAL

En los ejemplos expuestos hasta ahora hemos tenido en cuenta para nuestra decisión sólo uno de los dos posibles errores que podemos cometer con ella: controlamos y aceptamos una cierta probabilidad de clasificar como problemático algo que no lo es, pero no sabemos hasta qué punto con nuestra decisión estamos aceptando como bueno algo que en realidad no cumple el estándar. Es

como si en un test de screening conociésemos sólo su especificidad pero no su sensibilidad, o viceversa. En términos estadísticos, manejamos sólo uno de los dos errores posibles, α ó ß, Tipo I ó Tipo II. En términos del Control de Calidad en la industria, es como decidir sobre la base únicamente del riesgo del productor: la probabilidad de que un buen producto se rechace al clasificarlo como malo; pero sin tener en cuenta el riesgo del consumidor: probabilidad de que un mal producto se acepte como bueno. El LQAS (muestreo para la aceptación de lotes), basado también generalmente en la distribución binomial, contempla explícitamente ambos riesgos o tipos de error, una vez definido qué entendemos por "buen" producto (estándar de calidad del indicador) y "mal" producto (límite inaceptable o umbral del indicador que lo hace definitivamente problemático).

Para contrastar adecuadamente la hipótesis de si se cumple o no el estándar prefijado deberíamos considerar no sólo la probabilidad de que rechacemos que se cumple cuando en realidad sí se cumple, sino también la de aceptar que se cumple cuando en realidad no se cumple.

El LQAS, y sus tablas tal como se ha desarrollado en la industria, puede parecer complejo y difícil de manejar; sin embargo su simplificación y adaptación para la monitorización de indicadores de salud, es relativamente sencilla y puede limitarse a un manejo adecuado de las tablas de probabilidades acumuladas de la distribución binomial o, más simple aún, de las tablas elaboradas directamente para su uso con LQAS.

3. EL MUESTREO DE ACEPTACIÓN DE LOTES COMO MÉTODO DE MONITORIZACIÓN EN LOS SERVICIOS DE SALUD

El LQAS tiene en cuenta ambos tipos de posibles errores en la decisión que tomamos al contrastar la hipótesis de cumplimiento del estándar.

Tal como se define en la literatura sobre control de calidad en la industria, el muestreo para la aceptación de lotes es el proceso de evaluar una porción (muestra) de un lote de un determinado producto, con el propósito de aceptar o rechazar el lote en su totalidad. Las decisiones se toman sobre la base de la probabilidad de encontrar un número determinado de casos defectuosos (número de decisión), en muestras tomadas de cada lote (que sería el marco muestral), asumiendo la existencia en el lote de un determinado nivel de calidad o porcentaje de cumplimiento de los requisitos inspeccionados. El número de decisión se escoge de manera que los riesgos de rechazar lotes "buenos" (es decir, rechazar lotes que en realidad tengan el nivel de calidad deseado: riesgo alfa, o "riesgo del productor"), y de aceptar lotes "malos" (riesgo beta o "riesgo del consumidor") estén a un nivel conveniente y predeterminado. Estos riesgos y el plan de muestreo en sí mismo se basan en la distribución binomial en la mayoría de los esquemas, o en la de Poisson (si la proporción de casos defectuosos es muy baja, inferior al 1%). Existen también esquemas de LQAS para variables cuantitativas continuas, tomando entonces la decisión sobre la base de medias y desviaciones estándar, pero aquí tratamos sólo del LQAS para variables cualitativas dicotómicas (cumple/no cumple).

Para monitorizar con LQAS hay que especificar además del nivel de calidad que se quiere comprobar (estándar), los riesgos a y b que estemos dispuestos a correr: lo que equivale a especificar la probabilidad de falsos positivos y falsos negativos que queremos que tenga nuestro proceder para identificar situaciones problemáticas.

Los parámetros que definen el plan de muestreo son: (i) el tamaño del lote (o marco muestral), (ii) el tamaño de la muestra a evaluar, y (iii) el número para la aceptación o número decisional, una vez que hemos definido los estándares en relación al nivel de calidad asumido (llamados índices de calidad en la literatura de la industria), y los riesgos (alfa y beta) del muestreo.

- **El tamaño del lote** tiene poco efecto sobre la probabilidad de aceptación: en general, el tamaño del lote no importa a menos que el tamaño de la muestra sea mayor que el 10% del tamaño del lote. Sin embargo se aconseja que cuanto más grandes sean los lotes, más pequeños deben ser los riesgos o

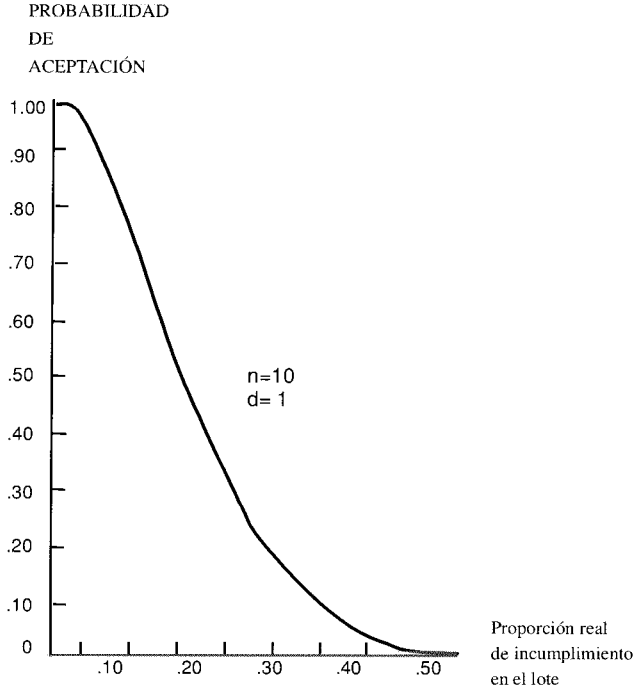
errores admitidos para la decisión, de forma que se minimicen las consecuencias de decisiones incorrectas. En los servicios de salud, conviene recordar que más que el tamaño de lote, que puede ser definido en términos de los casos ocurridos en un determinado periodo de tiempo (visitas en un año, diagnósticos en los últimos seis meses, etc.), hay que prestarle atención a la homogeneidad del lote: debe estar constituido por el mismo tipo de casos, igual que en la industria se inspeccionan lotes del mismo producto, y no mezclas de ellos.

Para aplicar el LQAS, más que el tamaño del "lote" (marco muestral) es importante considerar que este sea homogéneo (mismo tipo de paciente o casos a evaluar).

- **El tamaño de la muestra y el número decisional** dependen en gran medida de la definición de los índices de calidad y los riesgos que se acepta asumir con las decisiones. La combinación de tamaño de muestra y número decisional más eficientes para las características que queremos que tenga nuestro plan de monitorización, pueden encontrarse en tablas y gráficos (llamados Curvas de Características Operativas, Operating Characteristics Curves, OC) en la literatura de la industria. Tablas y gráficos no exentos de una cierta dificultad de comprensión y manejo, pero que podemos simplificar para nuestros propósitos entendiendo sus raíces y mecánica de construcción. La Figura 15.2 contiene el esquema básico de una curva OC; estas curvas representan la probabilidad de aceptación de diversos niveles de calidad, para un determinado tamaño de muestra y número de decisión. Hay, pues, una curva OC para cada número de decisión, una vez fijado el tamaño de la muestra. Son, esencialmente, la representación gráfica de las probabilidades binomiales acumuladas que corresponden a un determinado número decisional, en un determinado tamaño de muestra, dependiendo del valor real de incumplimientos que pueda tener el lote de donde se extrae la muestra.

El tamaño de la muestra para LQAS depende sobre todo de los índices de calidad prefijados para el indicador (estándar y umbral) y los riesgos de equivocarse (falsos positivos y falsos negativos) que se quieren asumir.

FIGURA 15.2. Ejemplo de curva de características operativas * (n=10, d=1)



* Hay una curva de características operativas para cada valor de "d", dado un determinado tamaño de muestra.
Fte.: Adaptado de Schilling EG.

Las curvas OC representan gráficamente la relación entre los diversos resultados posibles en una muestra y los riesgos de equivocarse al aceptar/rechazar un determinado nivel de calidad.

Una de las tablas más fáciles de manejar de entre las originadas en la industria es la ANSI/ASQL Z1.4, adaptación de la MIL-STD-105E (Military Standard 105E) y equivalente a la ISO 2859 (ANSI son las siglas de American National Standards Institute, e ISO son las siglas de International Organization for Standardization). Son en realidad un conjunto de tablas basadas en diversos valores de AQL y errores alfa y diversos esquemas, más o menos rígidos, de inspección. Su aspecto a primera vista es complejo. Sin embargo, existen algunas tablas que adoptan el LQAS a los servicios de salud, cuya comprensión y manejo son más sencillos que los originales de la industria. Dos de las pocas propuestas para uso en servicios de salud son las de Lemeshow et al. y la de JJ Valadez, para valorar esencialmente prevalencias y coberturas vacunales, cuyo enfoque y forma de utilización son ligeramente diferentes, aunque la mecánica de ambos sea finalmente la misma e idéntica a la que debemos seguir para utilizar las tablas que hemos adaptado para monitorización de indicadores de calidad y que incluimos en esta UT. Consiste en lo siguiente:

Las tablas de LQAS que se utilizan en la industria son complejas de entender, y proponemos aprender a manejar las que se han adaptado, simplificándolas, para su utilización en los servicios de salud.

1. Se define el "lote" o marco muestral en el que queremos comprobar el cumplimiento del indicador. Por ejemplo las visitas de un determinado tipo de pacientes en los últimos seis meses o un año, los pacientes incluidos en un determinado diagnóstico, programa o tratamiento, etc.

2. Se establecen el estándar (nivel de cumplimiento deseado, AQL) y el umbral (nivel mínimo de cumplimiento aceptable, LQL) para el indicador.

3. Decidimos los riesgos (alfa y beta, del "productor" y del "consumidor", ó falsos positivos y falsos negativos) que estamos dispuestos a tener al clasificar el "lote" como aceptable.

4. Buscamos en las tablas la combinación más eficiente (muestra más pequeña) de tamaño de muestra (n) y número decisional (d) con la que podemos efectuar la clasificación dado el estándar, el umbral y los riesgos prefijados.

Las tablas de LQAS indican el tamaño de muestra más eficiente y el número decisional (de cumplimientos o incumplimientos) para aceptar o rechazar si hay problema de calidad (incumplimiento o no del estándar prefijado).

5. Una vez identificados "n" y "d", la práctica de LQAS es tremendamente simple: se reduce a la extracción de una muestra aleatoria de "n" casos tomando el lote como marco muestral, y contar el número de casos en los que el requisito de calidad evaluado no se cumple. Si este número es mayor que "d" rechazamos el "lote": concluimos que AQL (el estándar) no se cumple porque la probabilidad del resultado obtenido es muy baja si AQL fuese cierto; hay, por tanto, un problema de calidad. Si por el contrario el número de incumplimientos es igual o menor que "d", aceptamos la calidad del "lote", equivalente a aceptar que el estándar se cumple, sabiendo además cual es la probabilidad de que estemos aceptando un "lote" con cumplimiento a nivel de LQL o peor. En definitiva, lo único que hacemos es contar el número de incumplimientos y ver si es mayor o menor que "d". No hace falta calcular nada. Una vez identificados "n" y "d", cualquier persona puede monitorizar, sin necesidad de saber como se ha llegado a ellos.

   Recordemos, sin embargo, que "d" va a ser número de cumplimientos o incumplimientos (en realidad se complementan), en función de que el estándar lo definamos a su vez como cumplimiento o incumplimiento.

5. LQAS EN SERVICIOS DE SALUD: DOS EJEMPLOS DE APLICACIÓN EN LA PRÁCTICA

Aunque las normas para decidir pueden derivarse de las tablas de la distribución binomial, proponemos para monitorización con LQAS utilizar las tablas ya elaboradas que se incluyen como Tablas 15.5 y 15.6 en esta UT, adaptadas de las confeccionadas por Lemeshow et al. para uso en servicios de salud.

Vamos a suponer que queremos monitorizar con LQAS los dos ejemplos de indicadores que hemos utilizado como ejemplo el principio de esta UT: la proporción de visitas a las que se les ha tomado la tensión arterial según las normas de detección que tengamos, y la proporción de pacientes diagnosticados de resfriado común a los que se les ha prescrito antibióticos de forma inadecuada, medidos ambos cada seis meses. Tal como lo hemos definido hasta ahora en estos ejemplos AQL (el estándar) es de 85% para la hipertensión y 15% para la prescripción de antibióticos (85% también, si lo definimos como no prescripción de antibióticos).

Necesitamos definir ahora LQL (el umbral) o nivel mínimo aceptable. Para la detección de hipertensos podemos establecerlo en 60%; sin embargo para la no prescripción de antibióticos, nos parece que deberíamos ser más exigentes y lo situamos en 65%; incluso consideramos que el estándar es deseable que lo situemos más alto: 90%.

A continuación establecemos que a la hora de identificar situaciones problemáticas nos parece aceptable un error alfa (falsos positivos) de 5% y una especificidad de 80% (20% de falsos negativos). Para buscar "n" y "d" tenemos tres opciones: podemos utilizar las tablas de probabilidad acumuladas de la distribución binomial, o las tablas para LQAS adaptadas de las de Lemeshow como las que se reproducen, en la Tablas 15.5 y 15.6, o las de Valadez.

TABLA 15.5. Tamaño de muestra (n) y número de decisión (c) para nivel de significación: 5%; poder: 80% (error α= 5%, error ß= 20%)

	ESTÁNDAR (P_0%)																					
Umbral	50		55		60		65		70		75		80		85		90		95	
(P_1%)	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c
10	8	1	6	1	5	1	*		*		*		*		*		*		*	
15	11	2	8	2	7	2	5	1	*		*		*		*		*		*	
20	15	4	11	3	9	2	7	2	5	1	*		*		*		*		*	
25	23	7	16	5	12	4	9	3	7	2	5	2	*		*		*		*	
30	37	13	24	9	16	6	12	5	9	4	6	2	5	2	*		*		*	
35	67	26	38	15	24	10	16	7	11	5	8	3	6	3	*		*		*	
40	153	66	68	30	38	17	23	11	16	8	11	5	8	4	5	2	*		*	
45	617	288	154	74	67	33	37	19	22	11	15	8	10	5	7	4	5	3	*	
50			615	317	151	80	65	35	35	20	21	12	13	8	9	5	6	4	*	
55					600	340	145	84	62	37	32	19	19	12	12	8	7	4	*	
60							573	353	136	86	57	37	29	19	16	11	10	7	5	3
65									534	356	125	85	50	35	25	18	13	9	7	5
70											483	346	109	80	43	32	20	15	9	7
75													419	321	91	71	33	26	14	11
80															342	279	69	58	22	19
85																	253	219	44	39

La Tabla 15.5 contiene los tamaños de muestra y número decisional para decisiones que arriesgan un 5% de falsos positivos y un 20% de falsos negativos en la identificación de situaciones problemáticas. La forma en que está construida la tabla incluye el tamaño de muestra más eficiente y el número decisional en el mismo sentido que el estándar; es decir, la decisión se toma sobre la existencia de un número mínimo de cumplimientos si el estándar es de cumplimiento, en vez de un número máximo de incumplimientos, como vimos en el apartado anterior. Esto es así porque Lemeshow et al. orientan su utilización para clasificar las poblaciones en función a niveles de prevalencia (de enfermedades, vacunación, etc.), de forma que lo que hemos llamado estándar de cumplimiento es para ellos nivel de prevalencia (que figura en la tabla como Po); el umbral o LQL, es en la tabla la prevalencia alternativa, Pa; lo que figura como "c" es el mínimo de casos de la enfermedad (o, en general, cumplimiento de lo que mida Po) que hay que encontrar en la muestra "n" para aceptar Po. Cualquier valor de cumplimientos menor a "c", nos ha de conducir a rechazar Po, es decir, rechazar la "calidad" del lote (que exista el nivel de prevalencia Po según el enfoque de Lemeshow).

Volviendo a nuestros ejemplos, la monitorización con LQAS del indicador sobre hipertensión (estándar ó Po =85%, umbral ó Pa =60%) podría hacerse, según la Tabla 15.5, con una muestra de 16 casos, en la cual debemos encontrar un mínimo de 11 cumplimientos. Para el indicador de la no prescripción de antibióticos en el resfriado común (estándar: 90%, umbral =65%), la monitorización la podemos hacer con una muestra de 13 casos, en la cual debemos encontrar un mínimo de 9 cumplimientos para no clasificar la situación como problemática. Estos valores de "n" y "c" se obtienen de la tabla, viendo donde se cruzan los valores asignados por nosotros a Po (estándar) y Pa (umbral). Así procederíamos para buscar los "n" y "c" más eficientes para cualquier estándar y umbral prefijados.

Puede observarse que cuanto más cerca esté Po de Pa, el tamaño de muestra que necesitamos es mayor. Por ejemplo, para un estándar de 85%, si el umbral lo ponemos en 65% (en vez de 60% como hemos hecho) necesitaríamos muestras de 25 casos. De igual manera si el umbral para la no prescripción de antibióticos lo elevamos a 70%, necesitaremos muestras de 20 casos para la monitorización; y 33 si lo elevamos a 75% (Tabla 15.5).

La muestra necesaria aumenta si al determinar las condiciones para monitorizar acercamos el umbral al estándar o disminuimos los riesgos de equivocarnos al aceptar/ rechazar si hay problema de calidad.

Por otra parte, si decidimos, disminuir los riesgos de equivocarnos al tomar la decisión (errores α y ß, también vamos a necesitar muestras más grandes. Por ejemplo, la Tabla 15.6 contiene los valores de "n" y "c" para que las decisiones de aceptar Po (estándar) tengan un error ß (aceptación de lotes con cumplimiento Pa), reducido al 10%. Así, utilizando los mismos ejemplos vemos que para monitorizar el indicador de hipertensión (Po =85%, Pa =60%) necesitaríamos una muestra de 24 casos (en vez de 16 que veíamos en la Tabla15.5); para la no prescripción de antibióticos (Po =90%, Pa =65%), necesitamos una muestra de 20 casos (en vez de los 13 que veíamos en la Tabla 15.5). Lemeshow et al. han construido tablas para diversos errores alfa y beta, con valores de error alfa de 1%, 5% y 10%; en combinación con errores beta de 10%, 20% y 50%. Con ello tenemos diversas opciones, de entre las cuales debemos escoger en función del tipo de problema a monitorizar y la importancia diferencial que podemos darle al riesgo de tener falsos positivos o falsos negativos, al decidir sobre la identificación de la situación como problemática.

La magnitud de los errores y que elijamos dependerá del efecto que pueda tener los falsos positivos y falsos negativos como resultado de la monitorización. Normalmente lo fijamos en 5% y en 20% ó 10%.

TABLA 15. 6. Tamaño de muestra (n) y número de decisión (c) para nivel de significación: 5%; poder: 90% (error α=5%; error ß=10%)

	ESTÁNDAR (P_0%)																					
Umbral	50		55		60		65		70		75		80		85		90		95	
(P_α%)	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c	n	c
10	10	2	8	2	6	1	5	1		*		*		*								
15	14	3	11	3	8	2	7	2	5	1		*		*		*						
20	20	6	15	5	11	3	9	3	7	2	5	2		*		*						
25	31	10	21	7	16	6	12	5	9	4	7	3	5	2		*		*				
30	50	19	32	12	22	9	16	7	12	5	9	4	7	3	5	2		*				
35	92	38	52	22	33	15	22	10	16	8	11	5	8	4	6	3	5	3		*		
40	211	93	93	43	52	25	32	16	22	11	15	8	11	6	8	5	6	4		*		
45	853	402	212	104	93	48	51	27	31	17	21	12	14	8	10	6	7	4		*		
50			852	444	210	114	91	51	49	29	30	18	19	12	13	8	9	6	5	3		
55					834	477	203	120	87	53	46	29	27	18	17	12	11	8	7	5		
60							798	496	191	123	80	53	42	29	24	17	14	10	8	6		
65									746	501	176	122	72	52	36	27	20	15	11	9		
70											676	488	156	116	62	48	30	24	15	12		
75													589	455	131	104	49	40	21	18		
80															484	398	102	86	34	30		
85																	362	316	67	60		

Las tablas construidas por Valadez, aunque tienen el mismo objetivo de servir para localizar el tamaño de muestra y el número decisional para LQAS, están elaboradas con un enfoque diferente. En estas tablas el punto de entrada es el tamaño de la muestra y el estándar de cumplimiento, valorando posteriormente qué nivel de LQL corresponde al error beta (falsos negativos) que estamos dispuestos a admitir. Para ello Valadez ha construido 45 tablas de probabilidad binomial, una para cada tamaño de muestra desde n=5 a n=50. En ellas se contemplan diversos niveles de cumplimiento del indicador a analizar (desde 5% hasta 95%), en las que el número decisional se da, sin embargo, como número de incumplimientos (recuérdese que tanto las tablas de probabilidad binomial con las de LQAS de Lemeshow et al. el estándar y el número decisional va en el mismo sentido). Con ello, una vez fijado el estándar, lo que se busca en la tabla es el tamaño de muestra más conveniente y el número máximo de incumplimientos que se pueden permitir para aceptar que la situación no es problemática (es decir, que el estándar se cumple), y los riesgos de equivocarse que conlleva según los diversos valores del umbral. Las tablas de Valadez son más numerosas y de manejo más engorroso que las de Lemeshow, por ello proponemos utilizar las tablas que hemos adaptado de Lemeshow y que incluimos en esta UT, al considerarlas suficientes para los planes de monitorización con LQAS en servicios de salud. Para una visión más amplia de todas ellas, puede consultarse la bibliografía original que se indica al final de esta UT.

El manejo de las tablas de monitorización con LQAS en la práctica es muy simple y no precisa de conocimientos estadísticos especiales ni de cálculo adicional alguno una vez fijados «n» y «c».

Una vez localizado en las tablas "n" y "c" según las condiciones (estándar, umbral, errores α y ß) que hemos decidido para la monitorización, ya no hace falta las tablas ni hacer ningún cálculo adicional para interpretar los resultados, hasta que decidamos cambiar los parámetros (por ejemplo el estándar) para la monitorización.

6. ESQUEMAS AVANZADOS DE MUESTREO DE ACEPTACIÓN DE LOTES

Es evidente que la muestra para monitorizar con LQAS puede dejar de ser pequeña si queremos acercar los valores de estándar y umbral o queremos reducir la probabilidad de equivocarnos en la decisión de aceptación/rechazo. Por este motivo, se han desarrollado en la industria diversos esquemas de LQAS que introducen modificaciones tendentes a minimizar este inconveniente y poder monitorizar con AQL y LQL relativamente cercanos y errores alfa y beta pequeños. Estas modificaciones incluyen (i) tablas con diversos niveles de rigor a utilizar de forma alternativa; (ii) esquemas de muestreos dobles y secuenciales; y (iii) esquemas de muestreo bayesianos.

6.1. DIFERENTES NIVELES DE EXIGENCIA SEGÚN LA EVOLUCIÓN Y EL COSTE DEL MUESTREO

Los métodos expuestos en los anteriores apartados son normas para aplicar a la inspección de lotes de forma independiente y teniendo en cuenta sólo los parámetros (AQL, LQL, riesgos α y ß) que hemos establecido. Sin embargo, si la monitorización se hace realmente rutinaria y con una cierta frecuencia, podemos replantearnos esquemas menos exigentes si lo normal viene siendo la no existencia de problemas de calidad, o, por el contrario, más estrictos, si nuestro esquema inicial viene detectando problemas con una cierta frecuencia.

Los esquemas de LQAS que combinan diversos niveles de exigencia, consideran explícitamente el coste de la inspección y el de la mala calidad que se detecta, establecen muestreos secuenciales, y se basan en probabilidad bayesiana, son modificaciones de gran interés pero aún no adaptadas para su uso en servicios de salud.

Cambios en los niveles de exigencia conllevan normalmente cambios en el tamaño de la muestra necesaria para decidir. Estos cambios están contemplados en la mayoría de las tablas para LQAS de la industria, que distinguen al menos tres niveles: inspección convencional, reducida y estricta. La reducida es la que emplea tamaños de muestra menores, y se aplica cuando los niveles de calidad vienen siendo consistentemente elevados, a la vez que se establecen señales de alarma para pasar a esquemas más estrictos, en función de la problemática que vayamos encontrando.

Otro aspecto que se considera para establecer el nivel de inspecciones es el coste de la inspección en sí, en relación a lo que puede evitar en términos de coste de mala calidad que pase inadvertida si no se inspecciona. Esta es, por ejemplo, la base de los tres niveles de inspección de las tablas ANSI/ASQC Z1.4 pensados para sistemas en los que la ratio entre el coste de inspección por caso (Ic) y el coste por caso de la mala calidad evitable por la inspección (Cn) son de 0,4, 1,0 y 1,6. En sus extremos, esta relación Ic/Cn=Pc puede llevar a aconsejar no inspeccionar o por el contrario revisar el 100% de los casos. En la decisión interviene la ratio de costes (Pc) y el nivel de calidad esperado (Qe) en términos de proporción de casos defectuosos (incumplimientos); en general la inspección será más estricta (hasta llegar al 100%, como en el caso de los indicadores

centinela) cuanto más pequeña sea la ratio (es decir, mayor el coste de la mala calidad en relación con el coste de la inspección) y mayor sea la diferencia entre el nivel de calidad esperada y el valor de la ratio de costes (Qe-Pc). Sin embargo, este tipo de cálculos, aunque interesantes, son relativamente más fáciles de hacer en la industria, con más tradición y delimitación clara de sus productos, y pueden resultar demasiados complejos para los servicios de salud.

6.2. LQAS CON MUESTREO DOBLE Y SECUENCIAL

Los muestreos secuenciales tratan de minimizar la realización de muestras ineficientes. Con este proceder se decide siempre con el mínimo tamaño de muestra posible, y, en el peor de los casos, la muestra total es equivalente a la que se realizaría para hacer una estimación del indicador.

Sobre la base de que las muestras pequeñas pueden detectar las situaciones claramente problemáticas, pero que conllevan errores mayores si la situación es intermedia, se han desarrollado en la industria tablas de LQAS con esquemas de muestreo doble, en las cuales se comienza con una muestra pequeña y número de decisión que permite detectar con pequeño riesgo de error situaciones muy problemáticas (LQL muy alejado de AQL), y se realiza una muestra adicional en el caso de que la primera más pequeña no permita decidir.

Dadas las ventajas de este proceder al evitar muestreos ineficientes (muestras más grandes de lo necesario para decidir), se han desarrollado con la misma filosofía esquemas de muestreo para LQAS múltiples o secuenciales. En ellos, se empieza con una muestra pequeña y el lote se acepta o se rechaza si el número de incumplimientos es muy pequeño o relativamente grande. Hay, pues, dos números decisionales, de forma que los resultados del muestreo con número de incumplimientos entre esos dos números decisionales, aplaza la decisión a una segunda muestra que se suma a la anterior, en donde se procede de igual manera; y así sucesivamente con una tercera, cuarta, etc. Siguiendo este proceder, podríamos llegar en el peor de los casos a tener una muestra total equivalente a la que necesitaríamos para una estimación del nivel de calidad con una determinada precisión en función de lo alejado que queramos que esté LQL del nivel AQL.

En el sector salud, el muestreo secuencial está siendo tímidamente empleado en algunos ensayos clínicos en los que se investiga la eficacia de un determinado tratamiento y puede resultar poco ético esperar a tener todos los casos de un muestreo tradicional, si puede decidirse antes con una cierta seguridad sobre la bondad o maldad del tratamiento. Sin embargo, el muestreo secuencial, a pesar de sus ventajas, al igual que el LQAS en general, sigue sin formar parte de los métodos de muestreo habituales en los servicios de salud.

6.3. MÉTODOS DE LQAS CON BASE BAYESIANA

En los procedimientos de LQAS vistos hasta aquí, se define el esquema de muestreo sobre unos parámetros predeterminados (AQL, LQL, errores α y β), pero no se tiene en cuenta el nivel real de calidad que puede existir en los lotes que se evalúan. Los planes de LQAS tradicionales están pensados para protegerse de un nivel de mala calidad determinado, que muchas veces no existe en la realidad que valoramos o tienen muy pocas probabilidades de existir; si esto es así, el resultado es que se utilizan muestreos demasiado grandes para protegernos (o identificar como problema) de niveles de mala calidad inexistentes. Con estas premisas, se han desarrollado tablas y planes de LQAS llamados planes de

muestreo empíricos o bayesianos que incorporan los datos de cada evaluación efectuada a la evaluación del lote siguiente. Comparaciones efectuadas entre los planes bayesianos y los habituales revelan que los bayesianos requieren muestras mucho más pequeñas, sobre todo cuando la variación en el tiempo de la proporción de incumplimientos es muy pequeña. El hecho de que haya que conocer (o asumir) un determinado nivel de calidad previo y que funcionen sobre todo cuando este nivel es estable, parece que ha conducido a que estos planes no se hayan utilizado más.

En su conjunto, sin embargo, el conocimiento y aplicación de los métodos de muestreo desarrollados en la industria para el control de la calidad, es un campo muy fecundo que puede hacer avanzar considerablemente la práctica de los programas de gestión de la calidad en los servicios de salud.

BIBLIOGRAFIA

- Richards LE, LaCava JJ. Business Statistics. Why and when. 2ª ed. New York: McGraw Hill; 1993.
- Schilling EG. Acceptance sampling in quality control. New York: Marcel Dekker Inc.; 1982.
- Lemeshow S, Hosmer DW, Klar J, Lwanga SK. Lot quality assurance sampling. En: Adequacy of sample size in health studies. Willshire: WHO/John Wiley & Sons; 1992.
- Valadez JJ. Assessing child survival programs in developing countries. Testing Lot Quality Assurance Sampling. Boston: Harvard University Press; 1991.
- Saturno PJ: Planes de Monitorización. Muestreo para la aceptación de Lotes. En: Saturno PJ, Gascón JJ, Parra P. Calidad Asistencial en Atención Primaria. Tomo II. Madrid: Dupont Pharma; 1997. p. 269-303.

16

LOS GRÁFICOS DE CONTROL ESTADÍSTICO DE LA CALIDAD: GRÁFICOS DE DESARROLLO

EMCA

Gestión de la Calidad Asistencial

CONTENIDO GENERAL

En esta UT se describen la base teórica del control estadístico de la Calidad y su aplicación en la monitorización y otras actividades de gestión de la calidad. Se presta una especial atención a los gráficos de desarrollo y de control, como herramientas de elección para el análisis y monitorización de indicadores con planes de mediciones frecuentes. Finalmente se ejemplifica el uso de los gráficos de desarrollo en los cuatro tipos de indicadores más habituales en servicios de salud.

ÍNDICE DE CONTENIDOS

OBJETIVOS ESPECÍFICOS

1. Entender la base teórica del control estadístico de la calidad.
2. Describir las características básicas de las herramientas de análisis estadístico gráfico.
3. Describir los objetivos esperables del análisis de datos con gráficos de control estadístico.
4. Distinguir los conceptos de variabilidad de causa común y variabilidad de causa especial.
5. Distinguir las diferentes situaciones en los que es de utilidad al control estadístico de la calidad.
6. Describir los requisitos de utilización del control estadístico para la monitorización continua de indicadores.
7. Distinguir la diferencia entre los gráficos x, ε, p y u.
8. Construir una plantilla de gráficos de desarrollo.
9. Interpretar los gráficos de desarrollo, identificando los patrones propios de variabilidad especial

1. INTRODUCCIÓN

El control estadístico de la calidad y sus herramientas más genuinas, los gráficos de control estadístico, son probablemente la aportación más importante del ámbito de la calidad en la industria para los programas de Gestión de la Calidad en cualquier actividad productiva o de servicios. Así, en la UT 14 se menciona el control estadístico (gráfico) de la calidad como el método de elección para monitorizar indicadores que se miden de forma continua o muy frecuente (como mínimo cada mes); a ello hay que añadir su utilidad como herramienta de análisis de los procesos, siendo posible 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 control estadístico de la calidad es el método de elección para analizar y monitorizar indicadores que se miden con una frecuencia mensual o menor. Son de utilidad en todas las actividades de gestión de la calidad

La utilización de los gráficos de control estadístico es simple, muy clara visualmente, y consiste esencialmente en un muestreo comparativo frecuente con muestras pequeñas. Para comprender como funcionan sólo hace falta un entendimiento básico sobre probabilidades y las distribuciones más frecuentes, sobre todo la normal; pero pueden ser utilizados aún sin estos conocimientos.

Tanto el análisis de la variabilidad como la monitorización del nivel estable de calidad, se realizan por medio de un uso masivo e inteligente de la teoría de la probabilidad, utilizando como referencia distintas funciones de probabilidad (normal, binomial, Poisson, etc.) según las características del indicador que se mide. Sin embargo, su utilización rutinaria está normalizada y no precisa de ningún conocimiento estadístico. Adicionalmente, el tamaño de muestra necesario para la monitorización puede ser pequeño (5 a 10 casos para variables cuantitativas, 25-30 para cualitativas).

El control estadístico (gráfico) de la calidad se basa en la teoría de la probabilidad y el uso de distintas funciones de probabilidad según el tipo de variable que mide el indicador.

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 estadística normal, o son estadísticamente incompatibles con una 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 análisis del control estadístico se realiza de forma visual, una vez construida la plantilla de referencia

La gran utilidad potencial del control gráfico de la calidad contrasta con lo lenta que está resultando su utilización rutinaria en los servicios de salud. Probablemente la principal causa es que se precisan series temporales con mediciones muy frecuentes para que puedan ser aplicados. Sin embargo, aparte de poder ser utilizados con éxito dentro de los ciclos de mejora y en el control de aspectos relevantes específicamente diseñados para su monitorización, en los servicios de salud se realizan muchas mediciones rutinarias con periodicidad mensual e incluso diaria cuyo análisis y presentación gráfica se beneficiarían enormemente si se hiciesen por medio de gráficos de control estadístico.

En esta UT vamos a describir las bases teóricas del control estadístico y ejemplificar los principales tipos de gráficos utilizables, centrándonos después en el más simple de todos: los gráficos de desarrollo o gráfico de rachas (run chart).

2. CONTROL ESTADÍSTICO DE LA CALIDAD. CONCEPTOS BÁSICOS

2.1. ¿QUÉ ES UN GRÁFICO PARA EL CONTROL ESTADÍSTICO DE LA CALIDAD?

Un gráfico para el control estadístico de la calidad es una plantilla que sirve de referencia para interpretar la variabilidad del indicador analizado

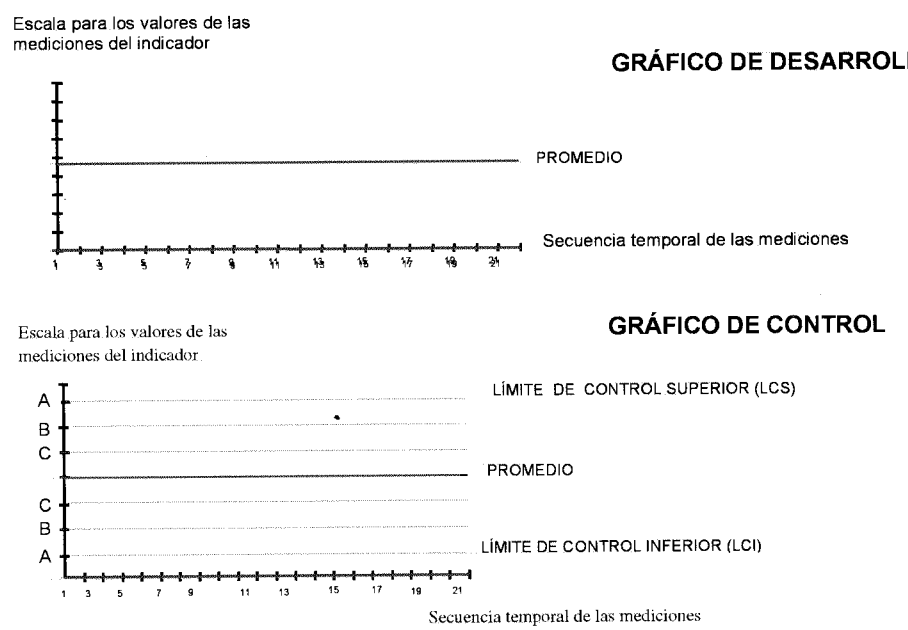
Los gráficos para el control estadístico de la calidad son esencialmente una plantilla en la que ir colocando las mediciones que se vayan haciendo del indicador a analizar o monitorizar. Sólo que esta plantilla está construida de una forma especial para que sirva de referencia en la interpretación de los valores del indicador a analizar o a controlar.

La propuesta original y utilización como herramientas para la identificación de fuentes de dificultades y defectos durante la producción la realizó Walter Shewart a principios de los años 20, pero su desarrollo y aplicaciones posteriores se han ido ampliando hasta constituir una de las ramas más fructíferas en cuanto a adaptaciones y utilización con ventaja sobre otros enfoques o métodos, en relación a prácticamente todas las actividades (ciclos de mejora, diseño, monitorización) de los programas de gestión de la calidad.

En la Figura 16.1 están representados los esquemas básicos de los dos grupos de gráficos de control estadístico que vamos a describir en detalle en esta UT y en la siguiente: gráfico de desarrollo y gráfico de control. En ambos casos la plantilla consta de un eje de ordenadas con la escala de valores que puede ir tomando el indicador, y un eje de abscisas con la secuencia temporal de las mediciones. Hay también en ambos una línea central, que representa el promedio del indicador a analizar y, en el caso del gráfico de control, dos líneas adicionales llamadas *límites de control* que delimitan la zona de variabilidad permisible para el indicador cuando el proceso que mide es estable. En estos ejes de coordenadas y con estas líneas de referencia, se van colocando los puntos correspondientes a las mediciones que vayamos realizando del indicador, pudiendo con ello analizar la variabilidad del proceso que medimos y concluir si es estable o no, en función de las posiciones que vayan teniendo los puntos sucesivos en el gráfico.

La plantilla de los gráficos más comunes consiste en unos ejes de coordenadas y una línea que representa el promedio estadístico del indicador (gráficos de desarrollo) y dos líneas adicionales (límites de control) que acotan la variabilidad esperable

FIGURA 16.1. Elementos de los gráficos para el control estadístico de la calidad



2.2. OBJETIVO PRINCIPAL: ANALIZAR LA VARIABILIDAD

El control estadístico de la calidad tiene como objetivo principal la monitorización continua de la estabilidad de los procesos de producción, sea de bienes o de servicios, de forma que se detecten las situaciones problemáticas ("fuera de control") y se pueda actuar sobre ellas. Para ello se realiza un *análisis gráfico* de la *variabilidad* de las mediciones del indicador para identificar lo que se conoce en la terminología del control de la calidad en la industria como causas "especiales", no aleatorias o "asignables" y las causas comunes, aleatorias o no asignables. Las causas "especiales" son aquellas que producen "inestabilidad", un "exceso" de variación no esperable por puro azar; las causas "comunes" son aquellas inherentes al proceso y que producen una variabilidad estable, sistemática y predecible, propia de la aleatoriedad de los elementos que intervienen en el proceso analizado. La distinción entre ambos tipos de causas es importante porque el tipo de investigación e intervención consiguiente también lo es. Las causas "especiales" son en cierto modo excepcionales y añadidas a lo que sería el funcionamiento normal del sistema, mientras que las "comunes" son inherentes al sistema en sí y una actuación sobre ellas significa una remodelación en profundidad, innovadora, del proceso analizado.

El control estadístico analiza la estabilidad de los procesos, distinguiendo variabilidad esperable por el azar, dentro del funcionamiento normal del proceso, de variabilidad no esperable. Las causas de la variabilidad esperable se les conoce como causas "comunes" y las de variabilidad no esperable o excesiva, causas "especiales".

Las decisiones sobre el tipo de variabilidad, especial o común, se basan en la teoría de la probabilidad, y más concretamente en las distribuciones de probabilidad que corresponden a cada tipo de indicador analizado. Estas decisiones son similares en su estructura a las pruebas de hipótesis, siendo la hipótesis nula (de base) el que el proceso monitorizado es estable, se mantiene a unos niveles constantes, y la variabilidad que se observa en las mediciones es propia del azar dentro de lo esperado para la variabilidad inherente al proceso en sí. La hipótesis alternativa sería que la variabilidad observada no es debida al azar, sino fruto de cambios en el proceso por causas que habría que averiguar.

Normalmente la actuación sobre causas comunes, para disminuir la variabilidad o cambiar el promedio, implica un rediseño de los procesos.

Cuando se monitoriza con gráficos de control estadístico, este contraste de hipótesis se realiza implícitamente con cada una de las mediciones del indicador, de forma que pueden identificarse de forma visual y rápida aquellas situaciones que son significativamente diferentes de la normalidad estadística (es decir, de un proceso estable con una variación en las mediciones debido a la variación normal esperable por azar). El riesgo de identificar como problemática una situación que no lo es (error α, Tipo I o falsos positivos) se ha establecido en la industria tradicionalmente en <0,01. La razón es puramente empírica, y busca minimizar investigaciones de causas especiales de forma innecesaria. El error ß, Tipo II o falsos negativos no es considerado explícitamente en las decisiones que se toman en los gráficos de control estadístico; si bien parece, siempre según la experiencia en la industria, que el proceder secuencial y centrado lo minimiza.

La identificación de uno y otro tipo de variabilidad se realiza en base a un contraste continuo de la hipótesis de estabilidad, según la distribución de probabilidades que corresponde al tipo de variable que mide el indicador.

2.3. CONDICIONES PARA SU UTILIZACIÓN COMO HERRAMIENTA DE MONITORIZACIÓN

Conviene señalar desde el principio que la monitorización con gráficos de control estadístico tiene sentido cuando se cumplen dos condiciones previas:

1. El proceso es estable o "estadísticamente controlado". Es decir, se observa con una serie de mediciones que los valores del indicador se mantienen dentro de la normalidad estadística (mientras siga habiendo una variabilidad

excesiva, indicativa de la existencia de causas "especiales", lo que hay que hacer es averiguarlas e intervenir sobre ellas).

2. Esta estabilidad o normalidad estadística está dentro de los límites de calidad que queremos para ese indicador (llamados especificaciones de calidad en el lenguaje de la industria).

La utilización del control estadístico como herramienta para la monitorización continua requiere que el indicador esté previamente estabilizado (variabilidad estadística predecible) y acorde con el nivel de calidad deseado.

En la consecución de estas dos condiciones también son útiles los gráficos de control estadístico, pero en este caso su uso es más como método para las diversas fases de un ciclo de mejora. La monitorización en sí como actividad de control (tal como la hemos definido en la UT 14) se haría una vez estabilizada la variabilidad y conseguidos los niveles de calidad que nos parecen aceptables.

Hay que señalar también que la consecución de estas dos condiciones debe ser secuencial, de modo que primero hay que conseguir la estabilidad del proceso y a continuación asegurarnos el cumplimiento ("conformidad") con los límites de calidad. En la Figura 16.2 están representados en forma de algoritmo la interrelación entre estas situaciones y su correspondencia con las actividades dentro del esquema de actividades de gestión de la calidad.

2.4. UTILIDAD DE LOS GRÁFICOS DE CONTROL ESTADÍSTICO

El hecho de que los gráficos de control estadístico puedan ser utilizados prácticamente en todas las actividades de los programas de gestión de la calidad, ha hecho que algunos autores los describan como el paradigma de la optimización sucesiva de los procesos de producción, un sistema en sí mismo para saber cómo lo estamos haciendo, de forma que sepamos cómo hacerlo mejor.

Además de para la monitorización continua, las herramientas del control estadístico sirven también para analizar la situación de partida de un indicador (identificación y análisis de problemas) comprobar el efecto de una intervención, y analizar la conformidad con los niveles de calidad requeridos

La versatilidad de usos ha llevado también a una cierta confusión entre los diversos objetivos y condiciones de aplicación para cada uno de ellos. Sin embargo, es muy importante no confundir objetivos y situaciones para los que se utiliza el control gráfico estadístico de los indicadores. Confusión que aparece incluso en algunos manuales de control de la calidad y que puede tener consecuencias prácticas indeseables. El hecho es que, aunque el método y las herramientas son similares, hay matices importantes según para qué se estén utilizando. Como resumen podemos decir que, los tres principales usos de los gráficos de control estadístico, cada uno con sus premisas y condiciones (Tabla 16.1), pueden ser:

Es conveniente tener claro en qué situación (ciclo de mejora, diseño, monitorización) y para qué objetivos vamos a utilizar las herramientas de control estadístico.

1. Análisis y control de la variabilidad de un proceso (representado a través de sus correspondientes indicadores). Aplicable a cualquier situación. Puede ser equivalente a la identificación y análisis de un problema de calidad u oportunidad de mejora. Su objetivo final es conseguir la estabilidad (control estadístico) del proceso, por medio de la identificación y posterior eliminación de las causas "especiales". Los gráficos de control sirven también, en este caso, para evidenciar el efecto de las intervenciones implementadas.
2. Análisis y control de la conformidad de *un proceso estable* con las especificaciones de calidad.
3. Monitorización de un proceso *estable y conforme* con las especificaciones, requisitos o estándares de calidad.

FIGURA 16.2. Algoritmo de utilización de los gráficos de control y su relación con las actividades de gestión de la calidad

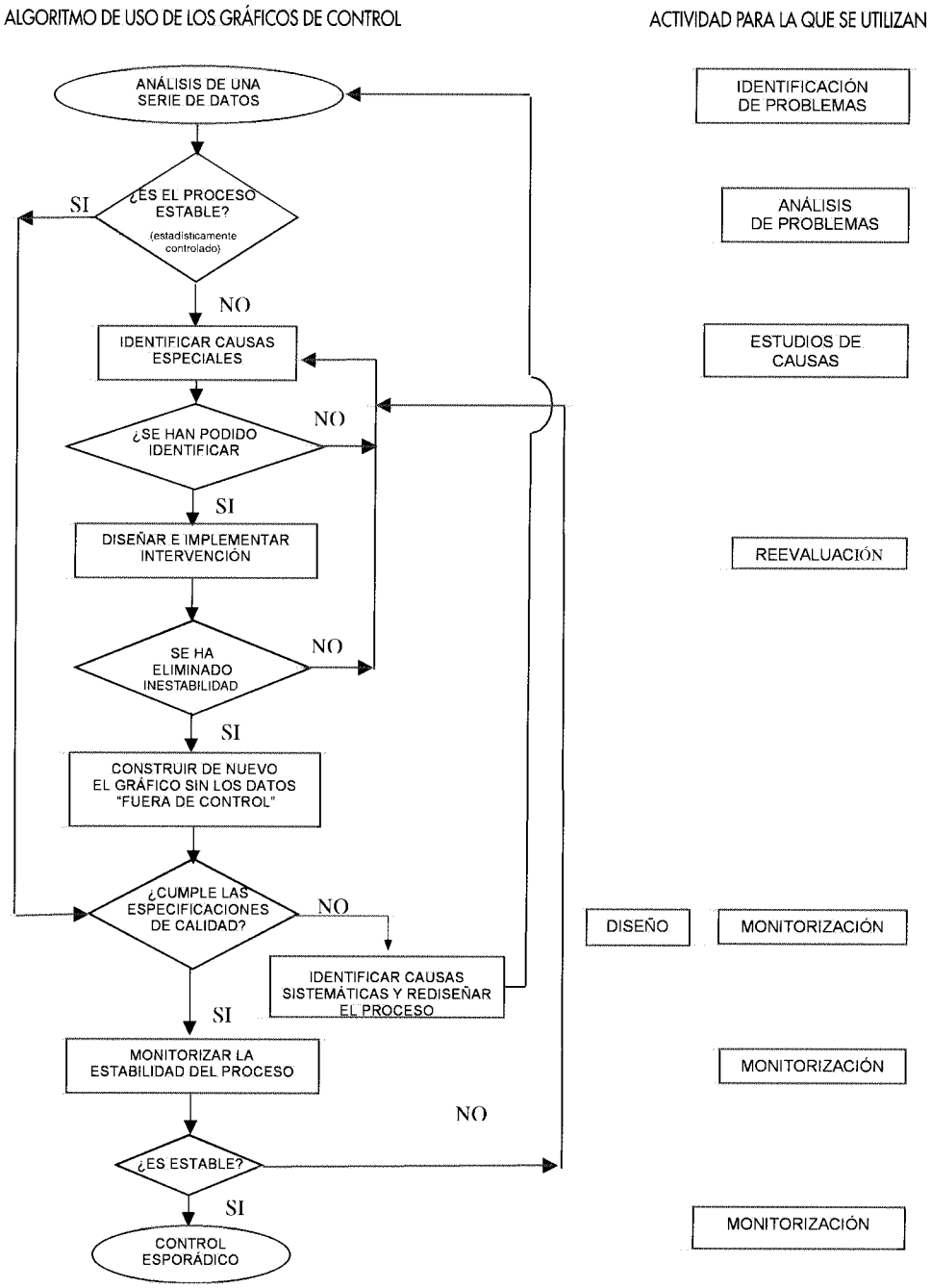


TABLA 16.1. Control estadístico (gráfico) de la calidad: utilidad, objetivos y lugar dentro de las actividades de gestión de la calidad

UTILIDAD	OBJETIVOS	TIPO DE ACTIVIDAD
ANÁLISIS DE SERIES DE DATOS (Análisis de procesos)	• ANÁLISIS DE VARIABILIDAD. • BÚSQUEDA DE CAUSAS • COMPROBAR EFECTO DE INTERVENCIÓN	CICLOS DE MEJORA (Identificación y análisis de problemas, estudio de causas, reevaluación)
ANÁLISIS DE LA "CAPACIDAD DE PROCESO" (En procesos estadísticamente estables)	• COMPROBAR SI LAS CARACTERÍSTICAS DE LA VARIABILIDAD INHERENTE AL PROCESO ANALIZADO AFECTAN A LOS REQUISITOS DE CALIDAD (ESPECIFICACIONES) QUE DEBE TENER	• DISEÑO DE LA CALIDAD • MONITORIZACIÓN
MONITORIZACIÓN (Control de procesos estadísticamente estables)	• CONTROL CONTINUO DE LA ESTABILIDAD DE UN DETERMINADO PROCESO (a través del indicador o indicadores oportunos) • IDENTIFICACIÓN DE SITUACIONES PROBLEMÁTICAS ("fuera de control") QUE HAY QUE INVESTIGAR.	MONITORIZACIÓN

3. LA BASE ESTADÍSTICA DE LOS GRÁFICOS DE CONTROL

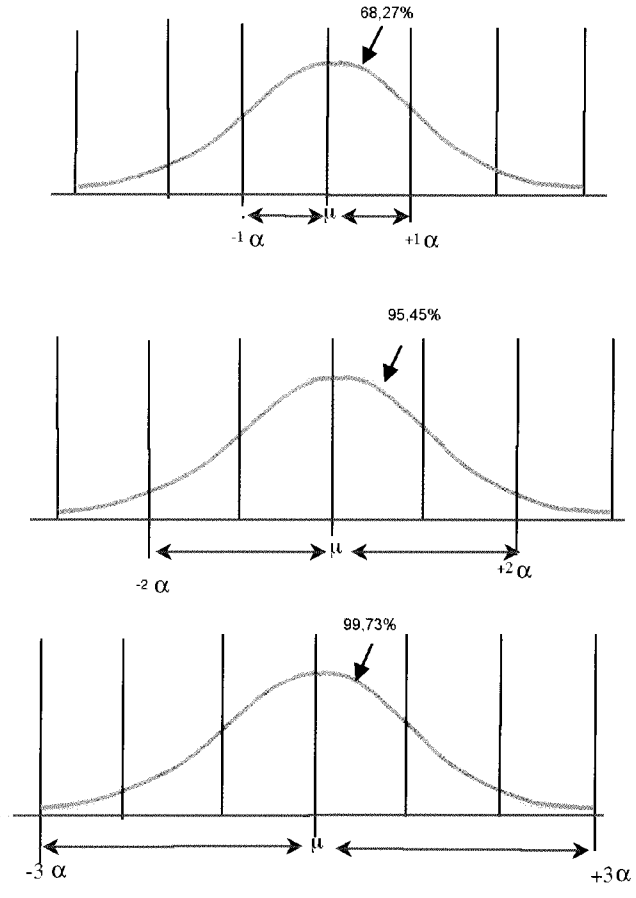
El razonamiento estadístico básico para el control gráfico de la calidad no puede ser más simple, ni su utilización práctica más ingeniosa. Se trata en esencia de lo siguiente:

En todo proceso de producción existe variabilidad, que puede ser debida a la variación en la forma de proceder (métodos de trabajo), el material que se emplee, la forma de funcionar de la tecnología empleada e incluso a la forma de medir el indicador para valorar la variabilidad. Ahora bien, si toda esta variación se mantiene dentro de unos límites normales y constantes, las mediciones que hagamos girarán en torno a un valor medio (el valor medio propio del proceso) siguiendo un patrón de probabilidades conocido y cuantificable (llamado en estadística *función* de probabilidad o *distribución* de probabilidad) que generalmente nos indica que los valores a encontrar en nuestras mediciones es más probable que resulten cercanos a la media propia del proceso que lejos de ella; de hecho cuanto más alejados estén del valor medio que caracteriza al proceso, menos probable será que nos aparezcan en nuestras mediciones, a menos que el proceso haya cambiado por alguna causa. Ese valor medio del indicador que medimos figura en el gráfico como línea central (Figura 16.1), con o sin otras líneas (límites de control) que delimitan los valores extremos esperables por azar. Lo que hacemos con cada medición del indicador es averiguar si *el valor que encontramos es o no compatible con la variación estadística normal en torno al valor medio.* Esta averiguación se hace visualmente (sin hacer cálculos) y se basa (como veremos después) no sólo en la medición que acabamos de realizar sino también en cuales han sido los resultados de mediciones anteriores (que también estarán en el gráfico) pero siempre considerando la probabilidad de obtener el valor que hemos obtenido si el valor medio permanece constante. De esta forma, si la probabilidad de que el valor que encontramos en la medición sea compatible con la situación de partida es muy baja (probabilidad que no hay que calcular sino que se deduce por la situación del punto que representa nuestra medición en el gráfico de control), concluiremos que algo "especial" ha ocurrido, el proceso está "fuera de control" y es conveniente averiguar cuál es la causa o causas de esta variación extraordinaria.

La base estadística del control gráfico de los indicadores es la probabilidad de que los valores en cada medición se alejen más o menos de un valor promedio. Si el proceso es estable, los valores de las mediciones es más probable que resulten cercanos a la media. Por otra parte, la probabilidad de cualquier valor viene determinada por una distribución estadística conocida de antemano.

Supongamos por ejemplo que la distribución de probabilidades que deben seguir las diversas mediciones del indicador si el proceso es estable es la distribución normal (la más conocida y la primera que explica cualquier libro de estadística elemental). Esta distribución se caracteriza por ser perfectamente simétrica en torno a la media, de forma que ésta ocupa el valor central y hay exactamente un 50% de probabilidades para valores por encima y 50% para valores por debajo, con probabilidades para cada valor individual tanto más pequeñas cuanto más lejanos estén del valor de la media. Probabilidades que son además perfectamente conocidas en función del número de desviaciones estándar por encima o por debajo de la media. En la Figura 16.3 están representadas estas probabilidades.

FIGURA 16.3. Distribución normal. Probabilidad de valores en torno a la media



Según las probabilidades de la distribución normal (Figura 16.3), si extraemos una muestra de un universo (marco muestral) de media μ, tenemos un 68,27% de probabilidades de que el valor que encontremos en la muestra esté comprendido en el intervalo determinado por la media ± 1 desviación estándar ($\mu \pm 1\sigma$); de igual manera hay un 95,45% de probabilidades de que esté comprendido en el intervalo $\mu \pm 2\sigma$ (es decir, menos del 0,05% de que encontremos en la muestra valores por encima o por debajo de $\mu \pm 2\sigma$); y 99,73% de que esté entre μ-3σ y

Si el proceso es estable, su valor medio es constante y el 99,7% de las mediciones ha de situarse entre el la media ± 3 desviaciones estándar.

$\mu + 3\sigma$. Quiere esto decir que si encontramos en nuestra medición valores fuera del intervalo $\mu \pm 3\sigma$ es muy poco probable (¡menos del 1%!) que el universo o marco muestral de donde hemos extraído la muestra, tenga la media μ que suponíamos debe tener. Concluiremos por tanto que ese nivel medio no se cumple, algo ha cambiado y la media del proceso ya no es la misma, se ha "desestabilizado".

Este razonamiento es aplicable al valor de una medición aislada del indicador, pero ya hemos apuntado que también se detectan situaciones de inestabilidad en base a la *secuencia de valores* que obtenemos en mediciones sucesivas del indicador; por ejemplo, es signo de inestabilidad si encontramos un determinado número de mediciones sucesivas a un mismo lado (por encima o por debajo) de la media, o constantemente hacia arriba o hacia abajo. ¿Por qué son estos patrones indicativos de cambios o alteración en el proceso analizado?: el razonamiento es igualmente probabilístico y tiene como base las probabilidades que seguirían las mediciones si se cumple la distribución normal (lo cual debe ocurrir en situaciones de estabilidad o control estadístico). En el caso de secuencias de valores nos interesaremos por la probabilidad no ya de un valor concreto (punto aislado del gráfico) como hemos hecho en el razonamiento anterior, sino de que tengamos una secuencia concreta de varios puntos (cada uno representa una medición del indicador) considerados conjuntamente. Esta probabilidad conjunta de toda una secuencia de valores se calcula, según nos indica la estadística, multiplicando las probabilidades de cada uno de los valores que estamos considerando.

Una determinada secuencia de valores en las mediciones de un indicador puede ser también un signo de inestabilidad estadística del proceso que se analiza o monitoriza.

Hagamos como ejemplo los cálculos para el caso más simple, con el objetivo de ilustrar cómo funciona, pero advirtiendo que en la práctica no es preciso hacer ningún cálculo porque están perfectamente definidos los patrones o secuencias de puntos en el gráfico que son estadísticamente incompatibles con que se mantenga la misma media; en la práctica sólo hay que ver si aparece alguno de estos patrones en el gráfico.

El caso más simple como decíamos, es considerar la probabilidad de una secuencia de valores del indicador a un mismo lado (por encima o por debajo) de la media. Sabemos que en la distribución normal hay un 50% de probabilidades de que una medición del indicador caiga a un lado determinado, y, por supuesto, un 100% de que caiga a uno u a otro lado indistintamente. Con estos supuestos, la probabilidad de que los valores del indicador que vamos obteniendo en mediciones sucesivas caigan siempre a un mismo lado de la media serían las siguientes:

Aunque los patrones de "descontrol" o "inestabilidad" problemáticos pueden calcularse en base a la función de probabilidades correspondiente, en la práctica no es necesario realizar calculo probabilístico alguno, el análisis es simplemente visual en busca de estos patrones que se conocen de antemano.

- Para la primera medición (primer punto a colocar en el gráfico): 0,5 (ó 50%)
- Para la segunda medición: 0,5 x 0,5= 0,25 (multiplicación de las dos probabilidades sucesivas)
- Para la tercera medición: 0,25 x 0,5= 0,125 (probabilidad anterior multiplicada por la probabilidad de un nuevo punto en el mismo lado) y así sucesivamente:
- Para la cuarta medición: 0,125 x 0,5= 0,0625.
- Para la quinta medición: 0,0625 x 0,5= 0,03125
- Para la sexta medición: 0,03125 x 0,5= 0,016

- Para la séptima medición: 0,016 x 0,5= 0,008

¿Qué quiere esto decir?: pues que la probabilidad de obtener cinco mediciones sucesivas todas ellas por encima o por debajo de la media es baja (<0,05; exactamente 0,03125) y debemos rechazar la hipótesis nula de que estamos ante la misma situación que teníamos antes de que ocurriese esta secuencia. Si queremos estar aún más seguros (disminuir el error α), esperaríamos a tener seis o siete mediciones sucesivas al mismo lado de la media; este último patrón tiene menos del 1% de probabilidades de darse, si la situación sigue siendo la que correspondería al promedio del gráfico.

Existen otros patrones o secuencias concretos de distribución de puntos en el gráfico que son indicativos, con una significación estadística determinada, de que la situación está "fuera de control", es decir que no corresponde a lo que sería esperable por azar si se mantiene en la realidad el mismo valor medio. Estos patrones los revisaremos más adelante; insistimos de nuevo en que en ninguno de los casos hay que hacer ningún cálculo en la práctica, ya que se han hecho previamente y se sabe que son patrones anormales en virtud a la distribución de probabilidades que deben seguir los valores del indicador si la situación es estable.

No todos los gráficos de control estadístico tienen como base la distribución normal, pero la forma de proceder y el razonamiento es idéntico en todos los casos: buscar el valor que represente la media general del indicador que medimos y ver la probabilidad de obtener los diversos valores, individualmente y en secuencia, que vayamos obteniendo con las mediciones sucesivas del indicador. En todos los casos, una vez construida la plantilla del gráfico, la decisión sobre la presencia o no de causas "especiales" (proceso "fuera de control", inestable, con media diferente a la que estamos utilizando en la monitorización) se toma por simple inspección visual del gráfico. En realidad, sea cual sea la distribución o función de probabilidad sobre la que se base el gráfico (normal, binomial, Poisson, no paramétrica), Shewart comprobó, y basó en ello las decisiones sobre la inestabilidad de los procesos, que se cumple la llamada desigualdad de Tchebyschev, según la cual la probabilidad de que la medición de una variable se desvíe de su media más de k veces su desviación estándar, es igual o menor que 1/k2, sea cual sea la distribución probabilística de base. Lo que cambia, al aplicar diversas distribuciones de probabilidad en el control gráfico, es la forma de calcular las desviaciones estándar y los valores promedio, pero la plantilla y la interpretación visual del gráfico son idénticas en todos los casos.

Sea cual sea la distribución de probabilidad que corresponda, el aspecto y la interpretación de los gráficos de control estadístico son idénticos. La única diferencia es la forma de calcular el promedio y los limites para la plantilla de referencia.

4. CONSTRUCCIÓN DE LA PLANTILLA DE REFERENCIA

¿Cómo se averigua el valor medio y los límites de referencia para construir la plantilla sobre la que graficar las mediciones y poder decir si la situación está o no "bajo control"? El conocimiento de estas características del proceso a monitorizar se realiza con una serie de mediciones sucesivas que utilizamos para estimar la media y desviación estándar, que nos sirven de base para construir la plantilla del gráfico y realizar los análisis o monitorizaciones posteriores. En estadística, las estimaciones de la media de una variable cuantitativa a través de muestra se anota como $\bar{x}$, y la desviación estándar como s. Veremos más adelante que, para el control gráfico, se utiliza para la construcción del promedio de la

El valor medio (línea central) y los limites de referencia se calculan y construyen a partir de al menos 20 mediciones sucesivas del indicador a analizar o monitorizar.

plantilla (y por tanto para tomar las decisiones sobre la estabilidad del indicador que se mide) la media de medias ($\bar{\bar{x}}$) o su equivalente según el tipo de variable que mide el indicador, y la desviación estándar de la media ($S\bar{x}$) para los límites de control.

¿Cuántas mediciones hacen falta para poder estimar $\bar{\bar{x}}$ y $S\bar{x}$? La mayoría de los manuales aconsejan al menos entre 20 y 30 muestras o mediciones sucesivas del indicador, a partir de las cuales construir la plantilla del gráfico. Esta es una de las razones por las cuales no sería posible utilizar los gráficos de control estadístico para indicadores que no se midan frecuentemente. Adicionalmente, para detectar situaciones fuera de control en la monitorización, si éstas aparecen no como valores extremos sino como secuencia atípica, nos haría falta en determinados gráficos un mínimo de cinco mediciones seguidas; ¿alguien se imagina tener que esperar cinco años o cinco semestres para detectar una situación problemática? Cuando Shewart propuso los gráficos de control estadístico como método de detectar problemas de la forma más inmediata posible en los procesos de producción, llegó a proponer, realizar y analizar determinados indicadores con una periodicidad de una hora entre mediciones. Sin llegar a estos extremos el control estadístico de la calidad puede aplicarse con provecho en los servicios de salud en mediciones que pueden ser hasta mensuales, tal como veremos más adelante.

5. ANÁLISIS GRÁFICO DE LA VARIABILIDAD: LOS CUATRO PRINCIPALES TIPOS DE GRÁFICO

Existe una gran cantidad de gráficos para el llamado Control Estadístico de Procesos, habitualmente clasificados según el tipo de variable que se analiza.

Aunque todos responden a un mismo esquema metodológico básico, existe gran cantidad de tipos diferentes de gráficos para el control estadístico de la calidad, y siguen apareciendo continuamente nuevas modalidades. De hecho el Control Estadístico de Procesos (Statistical Process Control, SPC, definido como la aplicación de métodos estadísticos en la medición y análisis de la variación) en el cual se encuadra y es principal protagonista el análisis gráfico, es prácticamente una disciplina en sí mismo, dentro del amplio campo de la gestión de la calidad. En algunos manuales de SPC se proponen una serie de clasificaciones para los diferentes tipos de gráfico como la que se recoge en la Tabla 16.2 pero quizás lo más útil en términos prácticos sea la consideración del tipo de gráficos no en función del tipo de variable, sino en función de una complejidad creciente en cuanto a su construcción sea cuál sea el tipo de variable que se mide, aparte de unas posibilidades de utilización también diferentes. Vamos a distinguir cuatro grupos que son los gráficos de *desarrollo*, los de *control*, los de *capacidad de proceso* y los de *cálculos secuenciales*.

TABLA 16.2. Gráficos de control estadístico de uso más frecuente en la industria. Clasificación según el tipo de dato que se mide

GRÁFICOS SOBRE INDICADORES QUE MIDEN VARIABLES CUANTITATIVAS	GRÁFICOS SOBRE INDICADORES QUE MIDEN VARIABLES CUALITATIVAS
• **Gráficos de valores medios** $\bar{x}$, gráfico de medias x, gráfico de valores individuales $\tilde{\chi}$, gráfico de medianas NLG, gráfico de límites estrechos (pre-control) CUSUM, gráfico de suma acumulada de medias. EWMA, gráfico de media móvil exponencialmente ponderada.	• **Gráficos sobre mediciones de una variable dicotómica** p, gráfico de proporciones np, gráfico de número de defectos CUSUM, gráfico de suma acumulada de desviaciones
• **Gráficos de medidas de dispersión** R, gráfico de rangos S, gráfico de dimensiones MR, gráfico de rangos móviles NLG, gráfico de límites estrechos (pre-control)	• **Gráficos sobre mediciones de más de una variable dicotómica** u, gráfico de medias de defectos por unidad de estudio c, gráfico de número de defectos CUSUM, gráfico de suma acumulada de desviaciones.

- **Gráficos de desarrollo.** También llamados de "rachas" o, en algunos países latinoamericanos, gráficos de "corridas", con una traducción literal pero confusa del ingles original (run chart).

  Estos gráficos tienen como único valor de referencia el valor medio del indicador. Son, por lo tanto, los más sencillos de construir e interpretar, pero su utilidad se reduce al análisis de tendencias y detecta cambios que se evidencian por medio de una determinada secuencia de valores (en un mismo lado de la media, continuamente ascendentes o descendentes, etc.). De ellos nos ocuparemos en esta UT.

- **Gráficos de control.** En los que además del promedio, como en los gráficos de desarrollo, están representados los llamados "límites de control" superior e inferior, (LCS y LCI), que señalan normalmente el promedio más 3 desviaciones estándar (LCS, límite de control superior) y el promedio menos 3 desviaciones estándar (LCI, límite de control inferior). Suelen establecerse habitualmente otras líneas de referencia marcando 1 y 2 desviaciones estándar por encima y por debajo del promedio; esta subdivisión facilita la aplicación de patrones adicionales de distribución de las mediciones en el gráfico que son significativas de "descontrol" (problemas), aumentando la sensibilidad del gráfico para la identificación de situaciones problemáticas. Los gráficos de control son el objeto de la UT 17.

- **Gráficos de capacidad de proceso.** Con ellos se analiza la relación entre la variabilidad estadística normal (controlada o estable) del indicador y las especificaciones que debe tener para estar al nivel de calidad que queremos que tenga. Para ello se hacen intervenir otros límites, llamados límites de especificación superior e inferior (LES y LEI), contra los que se compara la distribución de valores del indicador en el gráfico de control en situación estable,

En nuestro caso vamos a distinguir cuatro tipos principales de gráficos, en función de su complejidad y utilidad: gráficos de desarrollo (o rachas), gráficos de control, gráficos de capacidad de proceso y gráficos de cálculos secuenciales

controlada, (una vez eliminadas todas las causas especiales, no aleatorias o asignables). Esta comparación entre los límites de especificación y los valores del indicador dentro de los límites de control estadístico, se realiza en base a una serie de índices cuyos valores indican cual seria la mejor decisión a tomar para lograr que el proceso analizado, aparte de ser estadísticamente estable, produzca el nivel de calidad que se requiere. Los gráficos de capacidad de proceso son esenciales para las actividades de diseño de procesos.

- **Gráficos de control secuenciales**, estadísticamente más elaborados. Incluimos aquí un serie de gráficos, como los gráficos CUSUM (Cummulative Sum, suma acumulada) y EWMA (Exponentially Weighted Moving Average, media móvil exponencialmente ponderada), que incluyen para su aplicación una recalculación continua de los límites y el promedio en base a los valores previos del indicador (EWMA), siendo incluso la diferencia con valores anteriores lo que se lleva al gráfico, en vez de la medición del indicador en sí (CUSUM). Con ello se consigue al parecer, detectar desviaciones de patrones de estabilidad con mayor prontitud y exactitud.

Tanto los gráficos de desarrollo como los de control, reciben diversos nombres (x, , p, u) según el tipo de variable que miden y resumen. No hay sin embargo diferencias, ni en su aspecto, ni en las normas de interpretación.

6. CONSTRUCCIÓN E INTERPRETACIÓN DE LOS GRÁFICOS DE CONTROL ESTADÍSTICO MÁS SENCILLOS

Todo aspecto cuantificable y medido con frecuencia es susceptible de ser analizado y monitorizado con un gráfico de control estadístico. Adicionalmente, aunque se utilicen distintas distribuciones de probabilidad para establecer los parámetros de referencia, la construcción, visualización e interpretación del análisis gráfico es semejante en todos ellos. Por ello, los distintos gráficos que vamos a ver parecerían uno solo, si no indicamos cuál es la variable que se mide y cómo se mide. Vamos a detallar la construcción y uso de los *gráficos de desarrollo* (en esta UT) y los *gráficos de control* (en la siguiente UT) que responden a los cuatro principales tipos de mediciones o indicadores (números absolutos, medias, proporciones y ratios) que pueden realizarse en nuestro medio

Según se resume en la Tabla 16.3, los diferentes tipos de gráfico (diferentes por el tipo de medición que se representa, no por el aspecto del gráfico) pueden basarse en la distribución normal, binomial o incluso ninguna en concreto (no paramétrica), en función de que midamos medias, proporciones o números absolutos. De la misma manera, los cuatro tipos de gráficos de control seleccionados, que veremos en detalle en la UT 17, se basan en diferentes distribuciones de probabilidad, en función de las características del indicador al que se aplican. No hay en cambio, ninguna diferencia, ni en el aspecto, ni en las normas para su interpretación, ni tampoco, tal como veremos, en la forma de construirlos.

TABLA 16.3. Tipos de gráficos de control estadístico más comunes, indicadores para los que se utilizan y distribución de probabilidad en que se basan

TIPO DE GRÁFICO	TIPO DE INDICADOR	DISTRIBUCIÓN DE PROBABILIDADES
GRÁFICO DE DESARROLLO	MEDIA, PROPORCIÓN O NÚMERO ABSOLUTO DEL ASPECTO A MONITORIZAR	NORMAL, BINOMIAL O NO PARAMÉTRICA
GRÁFICO "X" DE VALORES INDIVIDUALES	NÚMERO DE EVENTOS POR UNIDAD DE TIEMPO (ejemplo: visitas/día)	NORMAL
GRÁFICO " " o de MEDIAS (acompañado o no de RANGOS o de DESVIACIONES ESTÁNDAR)	MEDIA DE UNA VARIABLE CUANTITATIVA POR UNIDAD DE ESTUDIO (Ejemplo: Media de minutos/visita)	NORMAL
GRÁFICO "p" o de PROPORCIONES	PROPORCIÓN DE UNA CUALIDAD (cumple/no cumple) EN UNA MUESTRA DE CASOS (ejemplo: proporción de visitas que esperaron más de 30')	BINOMIAL
GRÁFICO "u" o de RATIOS	RATIO DE UNA (O VARIAS) CUALIDADES POR UNIDAD DE ESTUDIO (ejemplo: Nº de factores de riesgo indagados/caso evaluado)	POISSON

Vamos a ilustrar los procedimientos de construcción e interpretación de los ráficos de control estadístico aplicándolos en cuatro situaciones diferentes:

En la primera de ellas, queremos controlar el tiempo que dedicamos a las consultas a demanda, no programadas. Hace tiempo que nos propusimos como objetivo de calidad el que nos dé tiempo por lo menos a saludar correctamente al paciente, que nos explicara por qué venía a la consulta y conducir una mínima entrevista clínica. Nuestro objetivo es una media de 10 minutos, pero sabemos que son varios los factores que pueden justificar más tiempo y también menos, de forma que queremos controlar hasta qué punto la variabilidad es estable y se mantiene en torno a los valores esperados. En la Tabla 16.4 están los resultados de una monitorización realizada en 25 días sucesivos. En cada día se ha medido el tiempo de consulta en 5 pacientes elegidos al azar, sin que se supiese a quién le tocaría. El indicador mide una variable cuantitativa (tiempo en minutos) y se expresa como media por paciente. ¿Cuál es nuestro promedio?, ¿se mantiene estable?, ¿debemos hacer algo para mejorar?

TABLA 16.4. Indicador: media de minutos por consulta a demanda. Resultado de 25 mediciones en días sucesivos

DÍA	TAMAÑO DE LA MUESTRA	MEDIA
1	5	10,2
2	5	8,4
3	5	7,6
4	5	7,0
5	5	8,0
6	5	10,2
7	5	7,4
8	5	8,0
9	5	10,8
10	5	6,0
11	5	9,2
12	5	8,2
13	5	6,0
14	5	10,8
15	5	11,4
16	5	5,2
17	5	8,6
18	5	10,0
19	5	6,6
20	5	9,3
21	5	9,1
22	5	8,2
23	5	8,1
24	5	8,0
25	5	7,9
TOTAL		210,2

Cálculos para la línea promedio según las fórmulas de la Tabla 16.8 para gráfico de medias:

$$\bar{\bar{X}} = \frac{210,2}{25} = 8,4$$

En ésta y en la siguiente UT utilizaremos como ejemplo cuatro tipos de indicadores diferentes que se corresponden con cuatro tipos de gráficos y distribuciones de probabilidad: media (gráfico , distribución normal); proporción (gráfico p, distribución binomial); valores absolutos (gráfico x, no paramétrico o normal); y ratio (gráfico u, distribución de Poisson).

- **En la segunda situación,** el aspecto que nos preocupa es el control de la espera excesiva, después de que hemos puesto en marcha sistemas organizativos de cita previa más ágiles y flexibles, pensados para que las visitas de quienes vienen a consultarnos sean fluidas, sin aglomeraciones y con poco tiempo de espera. Establecemos como indicador del funcionamiento de nuestro sistema la proporción de pacientes que tienen que esperar más de 30 minutos para ser atendidos. Hemos realizado la medida de este indicador durante un mes (un total de 22 mediciones) en una muestra diaria de 50 pacientes; toda persona que haya tenido que esperar más de 30' consideraremos que incumple nuestro indicador de calidad. Los resultados de estas mediciones están en la Tabla 16.5.

TABLA 16.5. Indicador: proporción de pacientes que tienen que esperar más de 30' antes de ser atendidos. Resultados de las mediciones efectuadas en todos los días de un mes

MUESTRA NÚMERO	TAMAÑO DE MUESTRA (N)	NÚMERO DE INCUMPLIMIENTOS	PROPORCIÓN DE INCUMPLIMIENTOS
1	50	12	0,24
2	50	15	0,30
3	50	8	0,16
4	50	10	0,20
5	50	4	0,08
6	50	7	0,14
7	50	16	0,32
8	50	9	0,18
9	50	14	0,28
10	50	10	0,20
11	50	5	0,10
12	50	6	0,12
13	50	17	0,34
14	50	12	0,24
15	50	22	0,44
16	50	8	0,16
17	50	10	0,20
18	50	5	0,10
19	50	13	0,26
20	50	11	0,22
21	50	20	0,40
22	50	18	0,36
TOTAL	1.100	252	

Cálculos para la línea promedio según las fórmulas de la Tabla 16.8 para gráfico de proporciones:

$$\bar{p} = \frac{252}{1.100} = 0,23$$

- **Nuestro tercer ejemplo** es la monitorización del funcionamiento de un programa nuevo que sabemos que no tienen todos los centros, pero que nosotros, que tenemos una relación estupenda con nuestro hospital de referencia, hemos logrado poner en marcha y estamos orgullosos de ello. Se trata de la realización en nuestro centro de intervenciones de cirugía menor, que normalmente irían a engrosar la lista de espera de los cirujanos. No son muchas las intervenciones que podemos realizar pero queremos mostrar que estamos utilizando de forma regular la formación recibida y el equipamiento adquirido expresamente para este programa. Medimos el número de intervenciones realizadas por semana (no parece posible un indicador tipo media o proporción). La Tabla 16.6 contiene los resultados de las últimas 20 semanas. ¿Cómo podemos analizarlos? ¿estamos actuando de forma regular y constante?, ¿cómo podemos monitorizar el funcionamiento del programa?

TABLA 16.6. Indicador: número de intervenciones por semana en un centro de salud. Resultados de mediones efectuadas en 20 semanas

SEMANAS	NÚMERO DE INTERVENCIONES
1	13
2	7
3	10
4	8
5	11
6	12
7	13
8	12
9	12
10	8
11	9
12	5
13	7
14	9
15	9
16	5
17	7
18	8
19	10
20	11
Total	186

Cálculos para la línea promedio según las fórmulas de la Tabla 16.8 para gráfico de valores absolutos:

$$\bar{x} = \frac{186}{20} = 9{,}3$$

- **El cuarto ejemplo** es más complicado. Tiene también que ver con controlar una mejora importante conseguida en la relación con los compañeros de otro centro, servicio o nivel asistencial. Nos hemos puesto de acuerdo en qué información necesitan de nosotros cuando le remitimos un paciente, y qué esperamos nosotros de ellos cuando nos los devuelven una vez cumplido su cometido. Tenemos claro el concepto de "cliente interno" y sabemos que ambos tenemos que estar atentos a las necesidades y expectativas del otro para ofrecer un servicio de calidad y que éste redunde en una mejor atención a nuestros pacientes. Nos hemos comprometido a informarles de cuatro puntos concretos cada vez que remitamos un paciente: datos de filiación, concisa descripción de antecedentes y exploraciones realizadas, sospecha diagnóstica y tratamientos prescritos. La falta de cualquiera de los puntos mencionados lo consideramos defecto de calidad igualmente importante, de forma que monitorizamos *el número de defectos por unidad de estudio*, siendo la unidad de estudio cada uno de los documentos interconsulta en que hemos evaluamos el cumplimiento de los cuatro criterios. Realizamos la medición de una muestra

7. GRÁFICOS DE DESARROLLO

Para el gráfico de desarrollo o de rachas lo único que hay que calcular es el promedio para ubicar la línea central de la plantilla.

Para la construcción de un gráfico de desarrollo lo único que se necesita es calcular la *línea central o promedio*, que es la que nos servirá de referencia para interpretar las mediciones del indicador.

Para calcular el *promedio* necesitamos unas 20 a 30 mediciones del indicador, las cuales se promedian de la siguiente manera, siguiendo las fórmulas indicadas en la Tabla 16.8, según midamos medias, proporciones, valores absolutos o ratios.

TABLA 16.8. Fórmulas para el cálculo de la línea central

GRÁFICO	LÍNEA CENTRAL (PROMEDIO)
X valores individuales	$X = \frac{\text{n° total de eventos en el periodo estudiado}}{\text{n° mediciones realizadas}}$
$\bar{X}$ (medias)	$\bar{\bar{X}}$ = Media de las medias de cada medición del indicador en la serie que analizamos o utilizamos para construir el gráfico = $\frac{\Sigma \bar{x}_i}{k}$
p (proporciones)	$\bar{p}$ = Proporción media de la serie de mediciones analizada o utilizada para construir el gráfico $\frac{\Sigma p_i \cdot n_i}{\Sigma n_i}$
u (ratios)	$\bar{U} = \frac{\text{Total ocurrencias de la cualidades que se miden}}{\text{Total de casos evaluados}} = \frac{\Sigma c_i}{\Sigma n_i}$

n_i : Tamaño de la muestra en cada medición.
k : n° de mediciones (muestras) realizadas para analizar el indicador o construir el gráfico.
c : Cumplimientos (o incumplimientos) de diversos criterios, evaluados todos ellos simultáneamente en cada unidad de estudio.
i : Secuencia ordinal de las mediciones efectuadas (1, 2, 3, ...i).

7.1. GRÁFICO DE DESARROLLO PARA MEDIAS

El promedio se calcula en base a 20-30 mediciones sucesivas del indicador, y según la fórmula que corresponda al tipo estadístico de indicador (media, proporción, valor absoluto o ratio).

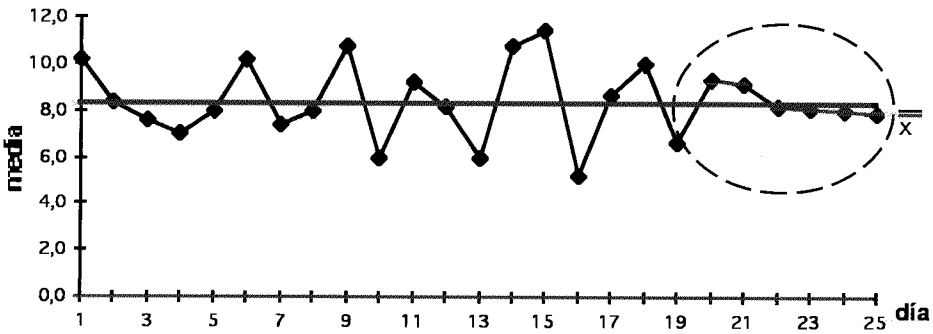
Si lo que medimos son *medias* (por ejemplo: la media de minutos por visita a demanda, el indicador de nuestro primer problema) se calcula la *media de medias* ($\bar{\bar{x}}$) de las 25 mediciones que hemos hecho del indicador; $\bar{\bar{x}}$ es la suma de la media obtenida (valor del indicador) en cada una de las mediciones, dividida por el número de mediciones (en nuestro caso 25), si todas las mediciones las hemos hecho con un mismo tamaño de muestra.

En la Tabla 16.4 está realizado este cálculo. Hemos encontrado que el promedio (la media de medias) de nuestra serie de mediciones es 8,4. A este nivel hemos de situar la línea central del gráfico para a continuación poner en él la secuencia de mediciones que hemos efectuado. El gráfico resultante puede verse en la Figura 16.4.

En caso de que el tamaño de muestra no fuese constante, el promedio habría que calcularlo de forma ponderada (multiplicando cada media por el tamaño de su muestra, sumar los resultados de estas multiplicaciones y dividirlo por el total

general de casos: $\Sigma (\bar{x} \cdot n_i) / \Sigma n_i$. Este sería el caso, por ejemplo, de haber medido el tiempo de consulta en todos los pacientes diarios, en vez de en una muestra, porque lo lógico es que este número total de cada día (equivalente en concepto a la muestra diaria) sea diferente cada día.

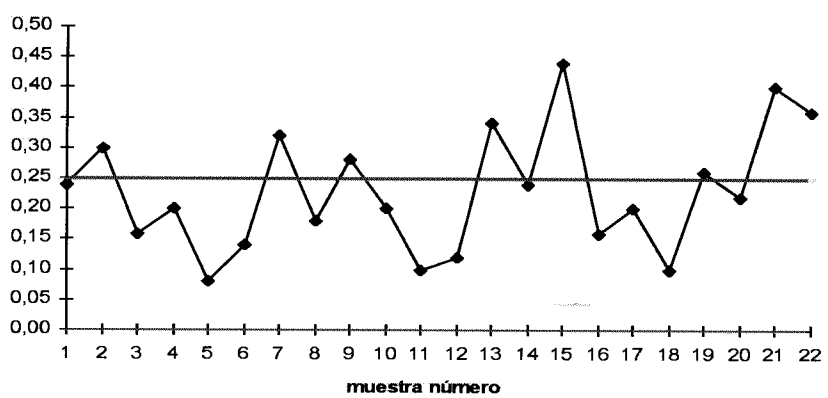
FIGURA 16.4. Media de minutos por consulta a demanda



7.2. GRÁFICO DE DESARROLLO PARA PROPORCIONES

Si lo que se mide son proporciones, calculamos igualmente la proporción media ($\bar{p}$) de todas las mediciones efectuadas. Este es el caso de nuestro segundo ejemplo en el que monitorizamos la proporción de pacientes que tienen que esperar en un determinado centro más de 30′ para ser atendidos, para lo cual hemos medido esta característica (esperar más de 30′ o no) en una muestra de 50 pacientes durante un mes (22 días). Como el tamaño de la muestra es constante, la proporción media sería la suma de todas las proporciones encontradas, dividido por el número de muestras, o, lo que resulta incluso más cómodo de calcular, el promedio *conjunto* (total general de incumplimientos dividido por total de casos evaluados). Si el tamaño de muestra no fuese siempre el mismo, habría que calcular siempre el promedio *conjunto* (considerar todas las muestras como si fuese una sola y contar el total de incumplimientos para dividirlo por el total de casos evaluados). En nuestro ejemplo (Tabla 16.5) tenemos un promedio de 23%. A este nivel ponemos la línea central, y sobre la plantilla así construida colocamos las 22 mediciones. El gráfico que resulta puede verse en la Figura 16.5.

FIGURA 16.5. Proporción de pacientes que tienen que esperar más de 30′

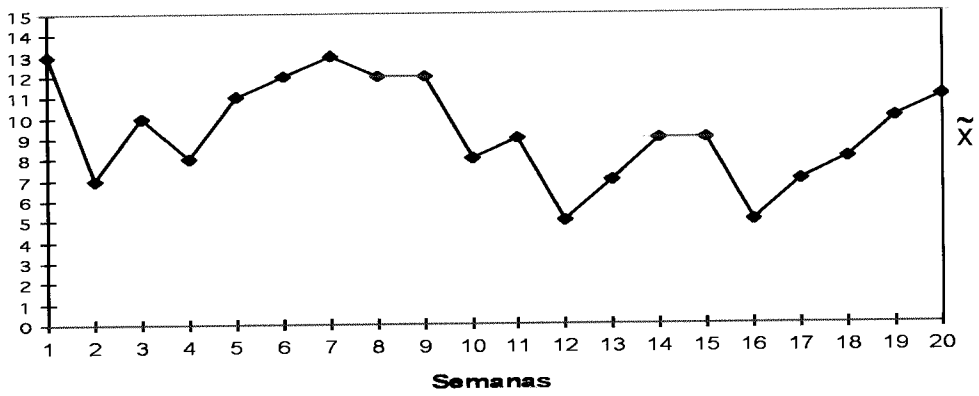


7.3. GRÁFICO DE DESARROLLO PARA VALORES ABSOLUTOS

Si nuestro indicador es el total de casos por cada unidad de tiempo en la que hacemos la medición, como en nuestro ejemplo del programa de cirugía menor (Tabla 16.6), calcularemos el promedio de todas las mediciones, o bien la mediana, lo cual sería más adecuado como referencia para interpretar los datos con test no paramétrico. La mediana ($\tilde{x}$) es el valor central de la serie, aquél que deja igual número de valores por encima y por debajo. Para hallarla hay que ordenar las mediciones desde el valor más pequeño (en nuestro ejemplo 5 intervenciones) al más grande (en nuestro ejemplo 13 intervenciones) y buscar el valor central. Si el número de mediciones fuese impar, la mediana sería simplemente el valor que ocupa el valor central directamente. Si la serie de mediciones es par (como en nuestro caso) la mediana es la media aritmética de los dos valores centrales. En nuestro caso los dos valores centrales de la serie ya ordenada de menor a mayor son los dos 9, de forma que la mediana es 9.

La línea central del gráfico sería entonces esta mediana de las 20 mediciones. El gráfico resultante de los datos puede verse en la Figura 16.6.

FIGURA 16.6. Número de intervenciones por semana

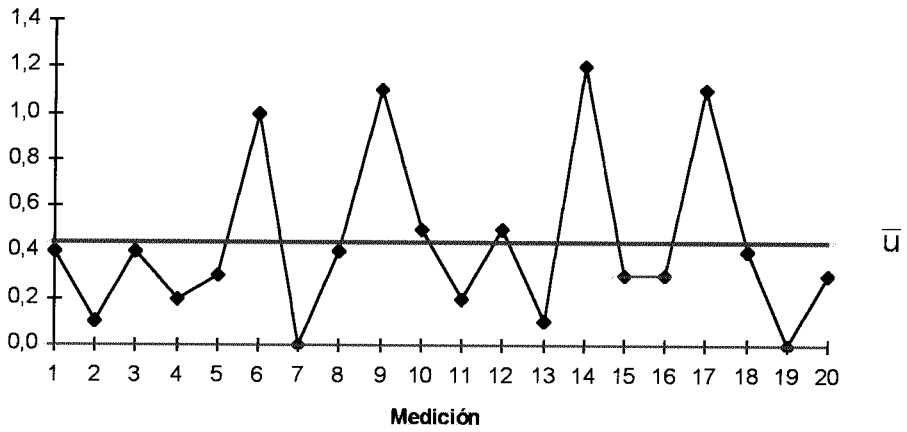


7.4. GRÁFICO DE DESARROLLO PARA RATIOS (Nº DE DEFECTOS POR UNIDAD DE ESTUDIO)

En nuestra monitorización de la calidad de la información contenida en el documento interconsulta (Tabla 16.7), el promedio para la línea central del gráfico lo calculamos como la media que resulta de sumar el número total de defectos encontrados y dividirla por el total de casos evaluados (suma de todas las muestras). Es, a efectos prácticos, como si fuese un gráfico de medias; si bien, la variable que se mide no es cuantitativa continua (como los minutos por visita) sino un conjunto de variables dicotómicas (cumple/no cumple) evaluadas todas a la vez en cada unidad de estudio. Ello hace que la distribución de probabilidades que vamos a utilizar de referencia para analizar los datos no sea la normal sino la de Poisson, pero *en la práctica* este matiz pasa inadvertido.

La misma situación, el mismo tipo de indicador, se da aunque sea sólo un criterio de calidad el que midamos, siempre que éste sea categórico (cumple/no cumple) y exista la posibilidad de que puedan aparecer más de un defecto (incumplimiento) en la unidad de estudio. Por ejemplo: números de errores tipográficos por hoja, número de palabras confusas o ilegibles por informe clínico, etc. El resultado gráfico de nuestro ejemplo esta en la Figura 16.7

FIGURA 16.7. Número de defectos por documento de interconsulta



8. INTERPRETACIÓN DE LOS GRÁFICOS DE DESARROLLO

Una vez calculado el promedio, y construida la "plantilla" del gráfico de desarrollo podemos *interpretar los resultados* de nuestras mediciones. Conviene recordar aquí que esta "plantilla" puede utilizarse tanto para analizar la variabilidad de los datos como para monitorizar futuras mediciones (esto último a condición de que no se aprecie en el análisis causas especiales de variación, y la estabilidad estadística incluya conformidad con las especificaciones de calidad).

En los gráficos de desarrollo se aprecian sobre todo tendencias significativamente diferentes de la situación de estabilidad.

En nuestros cuatro ejemplos vamos a utilizar el gráfico inicialmente para el primero de los tres usos que apuntábamos al comienzo de esta UT: *analizar la serie de datos* que hemos obtenido para cada indicador. Para ello hemos llevado al gráfico en su secuencia concreta todas las mediciones efectuadas con el resultado que aparece en las Figuras 16.4 a 16.7, y vamos a observar su distribución secuencial en busca de tendencias significativas. Los gráficos de desarrollo sirven básicamente para analizar y detectar tendencias.

Los patrones de inestabilidad o de diferencias significativas con el promedio lo constituyen una serie de puntos por encima o por debajo del promedio, en continuo ascenso o en continuo descenso.

Las secuencias o tendencias significativas de cambios o inestabilidad en el proceso son las que figuran en la Tabla 16.9. Obsérvese que son iguales para los indicadores que se expresan como media, proporción o ratio, mientras que para los números absolutos se toma la mediana como referencia y se aplican tests no paramétricos.

TABLA 16.9. Interpretación de los gráficos de desarrollo: patrones que denotan situaciones significativamente (p<0,01) diferentes de lo esperable por azar

<table>
<tr><th>TIPO DE INDICADOR</th><th>PATRÓN SIGNIFICATIVO DE VARIACIÓN</th></tr>
<tr><td>media, proporción, ratio</td><td>— 7 puntos consecutivos a un mismo lado (por encima o por debajo) del promedio.
— 6 puntos consecutivos en ascenso o en descenso.</td></tr>
<tr><td>números absolutos</td><td>Depende del número de mediciones (puntos) representados en el gráfico:
<table>
<tr><th>Nº mediciones</th><th>Puntos consecutivos un mismo lado de la mediana</th><th>Puntos consecutivos en ascenso o en descenso</th><th>Número mínimo de rachas* a un mismo lado de la mediana</th></tr>
<tr><td>20</td><td>8</td><td>7</td><td>4</td></tr>
<tr><td>30</td><td>9</td><td>7</td><td>8</td></tr>
<tr><td>40</td><td>10</td><td>7</td><td>12</td></tr>
<tr><td>50</td><td>11</td><td>7</td><td>16</td></tr>
</table></td></tr>
</table>

*: racha: conjunto de puntos consecutivos a un mismo lado de la mediana.
Fte: Elaboración propia a partir de Pyzdek T y Farnum.

Si observamos los gráficos (Figuras 16.4 a 16.7) buscando la existencia de alguno de estos patrones, se evidencia una situación problemática en el indicador relativo a la media de minutos por consulta: los últimos seis puntos en constante descenso; esta situación es clara y significantemente diferente del promedio y merece que investiguemos qué está pasando. La aparición de un patrón anormal se suele marcar con una cruz (x) en el gráfico.

En los otros gráficos no se observa ningún patrón anormal de los que figuran en la Tabla 16.9. Sin embargo, merece la pena que nos detengamos en la aplicación del *test de rachas* en el gráfico de números absolutos (Figura 16.6), porque no es tan evidente como los otros tests. Hay que realizar los siguientes pasos:

1. Contar el número de puntos por encima y por debajo de la mediana, y los que están justo en la mediana. En nuestro gráfico hay 9 puntos por encima, 8 puntos por debajo y 3 en la mediana.

   Los puntos que están en la mediana hay que asignarlos al grupo de los de arriba y los de abajo, de forma que quede un número igual de puntos a cada lado. En nuestro ejemplo habría que asignar 1 a los de arriba y 2 a los de abajo, de forma que queden 10 y 10. Si el número de puntos fuese impar, se descartaría para el test de rachas uno de los puntos situado justo en la mediana.

2. Se cuentan el número de "rachas" (puntos seguidos a un mismo lado de la mediana) que hay por encima y por debajo de la mediana en todo el gráfico. Los puntos que están justo en la mediana y que hemos asignado se escogen de uno u otro lado de forma que se maximice el número de rachas (si está a continuación de una racha de lado de abajo, se asigna a los de arriba y viceversa). Las rachas pueden ser de sólo un punto. En nuestro gráfico hay 5 rachas en el lado de arriba y 4 en el lado de abajo, contando con la asignación de los puntos que están en la mediana.

Cuando el indicador es de números absolutos, aunque puede calcularse la media, lo más apropiado es utilizar como promedio la mediana, y buscar patrones significativos de inestabilidad por medio del test de rachas, de naturaleza no paramétrica.

3. Se mira en las tablas (Tabla 16.9) si el número de rachas más pequeño de los dos (en nuestro caso las 4 rachas del lado de abajo) es igual o mayor que el mínimo esperable. Si resulta *menos* que el mínimo esperable, es un patrón que indica inestabilidad. En nuestro caso podemos ver que 4 rachas es precisamente el número mínimo esperable para un gráfico con 20 puntos, por lo tanto concluiremos que la evolución de este indicador está dentro de la variabilidad esperable por el azar en un proceso estable.

Cuando se utilizan los gráficos de desarrollo hay que procurar interpretarlos en su justa medida: esencialmente no dar significación a puntos aislados ni a patrones diferentes de los previamente conocidos como significativos

Realizar el análisis gráfico con gráficos de desarrollo tiene la ventaja de la simplicidad, pero también tiene otras desventajas y peligros:

- **Los peligros** existentes son: (i) que intentemos darle importancia a los "picos" en el gráfico, cuando sólo sirve para detectar tendencias; y (ii) que no nos terminemos de creer que las tendencias existen y son significativas y esperamos a tener "un punto más" para confirmarlo; en este caso, como la nueva situación también tendrá su propia variación aleatoria pueden aparecer mediciones que nos produzcan el espejismo de haber vuelto a la normalidad, y el problema queda sin identificar y puede agravarse y/o repetirse en el futuro.

- **La desventaja** es que es bastante menos sensible que los gráficos con límites de control tanto para analizar la variabilidad de los datos como para monitorizar cambios. De hecho hay manuales que no incluyen los gráficos de desarrollo entre el arsenal de gráficos para el control estadístico de la calidad. Sin embargo, su aplicación rutinaria, automatizada, a todos estos indicadores rutinarios y "obligatorios" que se miden en nuestro sistema de salud sería tremendamente útil o, cuando menos, mucho más informativo que los listados de números y que otras representaciones gráficas más habituales.

Los gráficos de desarrollo son muy sencillos de construir e interpretar, pero son menos sensibles que los gráficos de control para detectar situaciones problemáticas.

BIBLIOGRAFIA

- Deming WE. Calidad, productividad y competitividad. La salida de la crisis. Madrid: Díaz de Santos; 1989.
- Walpole RE, Myers RH. Probabilidad y estadística, 4ª ed. México: McGraw Hill; 1992.
- Plsek PE. Tutorial: Introduction to control charts. Quality management in healt care 1992, 1(1): 65-74.
- Ott ER, Schilling EG. Ideas from time secuences of observations. En: Process quality control. 2ª Ed. New York: McGraw Hill; 1990.
- Farnum NR: Control chart concepts. En: Modern Statistical quality control and improvement. Belmont: Duxbury Press; 1994.
- Wadsworth HM, Stephens KS, Godpey AB. Statistical process control. En: Modern methods for quality control and improvement. New York: John Wiley and sons; 1986.
- Manual de herramientas básicas para el análisis de datos. GOAL/QPC, Methuen, 1990.

17

LOS GRÁFICOS DE CONTROL ESTADÍSTICO DE LA CALIDAD: GRÁFICOS DE CONTROL

EMCA

Gestión de la Calidad Asistencial

CONTENIDO GENERAL

En esta UT se describen las características de los Gráficos de Control, incluyendo la forma de construirlos según los distintos tipos de indicadores (media, proporción, ratio, valores absolutos) y su interpretación. Se muestra cómo y por qué son más sensibles que los gráficos de desarrollo para detectar situaciones problemáticas y se advierte de los principales problemas metodológicos que hay que tener en cuenta para su utilización.

ÍNDICE DE CONTENIDOS

1. Introducción.
2. Gráficos de control: construcción de la plantilla gráfica para el análisis y monitorización de datos.
3. Interpretación de los gráficos de control.
4. Principales precauciones a tener en cuenta para el uso del control estadístico de la calidad.

OBJETIVOS ESPECÍFICOS

1. Describir las características básicas de los Gráficos de Control.
2. Construir una plantilla de Gráficos de Control.
3. Identificar en un Gráfico de Control los patrones propios de variabilidad especial.
4. Determinar el tamaño de muestra necesario para realizar mediciones a analizar con Gráficos de Control.
5. Tener en cuenta los principales problema que pueden conducir a interpretación errónea de los resultados.

1. INTRODUCCIÓN

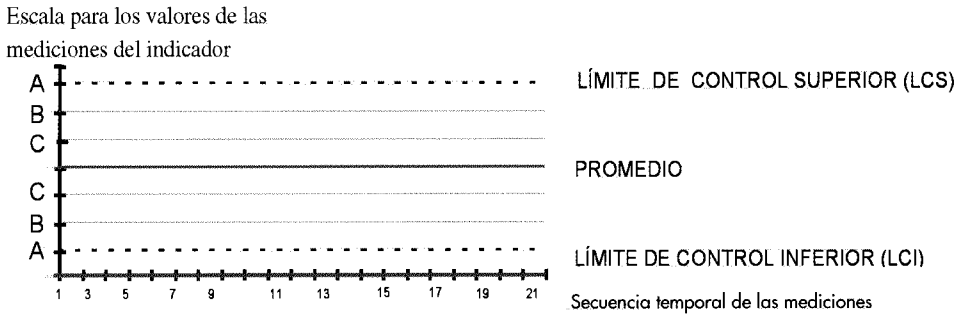
En la UT 16 hemos visto las bases conceptuales del control estadístico de la calidad, haciendo referencia desde el punto de vista práctico a los dos formatos más sencillos (gráfico de desarrollo y gráfico de control) y a los cuatro tipos de gráficos (x, $\bar{x}$, *p* y *u*) que se corresponden con los cuatro tipos más comunes de indicadores que pueden ser utilizados en los servicios de salud. Hemos visto también los ejemplos prácticos correspondientes en cuanto a indicadores, construcción de la plantilla de referencia e interpretación de los gráficos, en relación a los gráficos de desarrollo; esta ejemplificación continúa en esta UT, en la que analizaremos los mismos datos de los mismos cuatro tipos de indicadores, mediante gráficos de control.

2. GRÁFICOS DE CONTROL: CONSTRUCCIÓN DE LA PLANTILLA GRÁFICA PARA EL ANÁLISIS Y MONITORIZACIÓN DE DATOS

En relación a los gráficos de desarrollo que vimos en la UT 16, los gráficos de control incorporan, al menos, dos líneas de referencia más (los límites de control) que representan el promedio ±3 desviaciones estándar. De esta forma, la probabilidad de que una medición del indicador resulte fuera de los límites de control es <0,01. La plantilla de base es la representada en la Figura 17.1.

La plantilla de los gráficos de control consta del promedio y, al menos, dos límites de control que corresponden al promedio ± 3 desviaciones estándar

FIGURA 17.1. Gráfico de control



Las fórmulas para su cálculo son diferentes según el tipo de indicador desde el punto de vista estadístico. Las que corresponden o los cuatro tipos que hemos seleccionado figuran en la Tabla 17.1, y se aplican según vemos a continuación, utilizando los mismos ejemplos que hemos introducido en la UT 16.

TABLA 17.1. fórmulas para el cálculo de la línea central y límites de control de los gráficos más comunes

GRÁFICO	LÍNEA CENTRAL (PROMEDIO)	LÍMITES DE CONTROL
X valores individuales	$\bar{x} = \frac{\text{nº total de eventos en el periodo estudiado}}{\text{nº mediciones realizadas}}$	$\bar{x} \pm 3 \frac{\overline{RM}}{1{,}128}$
$\bar{x}$ (medias)	$\bar{\bar{X}}$ = Media de las medias de cada medición del indicador en la serie que analizamos o utilizamos para construir el gráfico = $\frac{\sum \bar{x}_i}{k}$	$\bar{\bar{X}} \pm 3$ Desv.Estándar de la media ó $\bar{\bar{X}} \pm A_2 \bar{R}$
p (proporciones)	$\bar{p}$ = Proporción media de la serie de mediciones analizada o utilizada para construir el gráfico = $\frac{\sum p_i \cdot n_i}{\sum n_i}$	$\bar{p} \pm 3 \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$ (*)
u (ratios)	$\bar{u} = \frac{\text{Total ocurrencias de las cualidades que se miden}}{\text{Total de casos evaluados}} = \frac{c_i}{n_i}$	$\bar{u} \pm 3 \sqrt{\frac{\bar{U}}{n_i}}$ *

$\overline{RM}$: Rango móvil retrospectivo medio (ver explicación en el texto).
$\bar{R}$: Rango medio de las mediciones efectuadas:
A_2: Una constante dependiente del tamaño de muestra empleado (ver tabla).
n_i : Tamaño de la muestra en cada medición.
k : nº de mediciones (muestras) realizadas para analizar el indicador o construir el gráfico.
c : Cumplimientos (o incumplimientos) de diversos criterios, evaluados todos ellos simultáneamente en cada unidad de estudio.
i : Secuencia ordinal de las mediciones efectuadas (1, 2, 3, ...i).
(*): Si el tamaño de la muestra no es constante, los límites son variables, según los diversos tamaños de muestra empleados. Si el tamaño de la muestra es constante, los gráficos "***np***" (número absoluto de cumplimiento o incumplimiento de un criterio o indicador) y "***c***" (número absoluto de cumplimientos o incumplimientos de varios criterios o indicadores evaluados simultáneamente en cada unidad de estudio), son otra posible alternativa.

2.1. GRÁFICO DE MEDIAS ($\bar{x}$)

Los límites de control del gráfico $\bar{x}$ pueden establecerse calculando la desviación estándar del promedio ($S\bar{x}$), o utilizando formulas que incluyen el rango medio

Hay dos maneras de proceder para establecer los límites de control, una es calculando la desviación estándar de la media ($S\bar{x}$), lo que equivale a calcular la desviación estándar para el conjunto de muestras como si fuese una sola y dividirlo por $\sqrt{n}$. La segunda manera es más cómoda y más común, se utiliza el rango medio ($\bar{R}$) multiplicado por una constante que varía con el tamaño de la muestra empleada cada vez que hemos medido el indicador, y que se busca en unas tablas como las que se reproducen en la Tabla 17.6. Esta tabla contiene otros factores que se utilizan para calcular los límites de control cuando el promedio es sobre el rango ($\bar{R}$) o sobre la desviación estándar ($\bar{S}$).

En la UT 16 vimos un ejemplo de indicador medido como media: media de minutos por consulta no programada; utilizando estos mismos datos que reproducimos aquí en la Tabla 17.2, el valor de A_2 en la Tabla 17.6 es 0,577, que es el que corresponde a las mediciones con muestras de 5 casos como las que hemos utilizado. Los límites de control quedan así establecidos en 11,7 (LCS) y 5,1 (LCI). El gráfico que resulta, tras poner en él los puntos que corresponden a las 25 mediciones hechas, puede verse en la Figura 17.2.

TABLA 17.2. Indicador: media de minutos por consulta a demanda. Resultado de 25 mediciones en días sucesivos

Día	Tamaño de la muestra	Valor Máximo	Valor Mínimo	Rango	Media
1	5	20'	3'	17	10,2
2	5	10'	5'	5	8,4
3	5	12'	5'	7	7,6
4	5	8'	6'	2	7,0
5	5	10'	7'	3	8,0
6	5	15'	5'	10	10,2
7	5	9'	5'	4	7,4
8	5	10'	6'	4	8,0
9	5	16'	8'	8	10,8
10	5	7'	5'	2	6,0
11	5	13'	6'	7	9,2
12	5	10'	5'	5	8,2
13	5	7'	5'	2	6,0
14	5	14'	7'	7	10,8
15	5	15'	6'	9	11,4
16	5	6'	5'	1	5,2
17	5	10'	7'	3	8,6
18	5	15'	6'	9	10,0
19	5	8'	5'	3	6,6
20	5	10'	7'	3	9,3
21	5	13'	6'	7	9,1
22	5	12'	5'	7	8,2
23	5	15'	6'	9	8,1
24	5	10'	5'	5	8,0
25	5	9'	3'	6	7,9
TOTAL				**145**	**210,2**

Cálculos para la línea promedio según las fórmulas de la Tabla 17.1 para gráfico de medias:

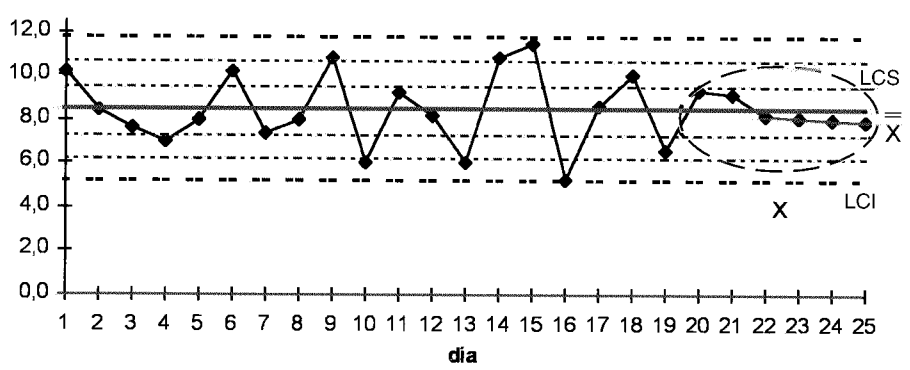
$$\bar{\bar{X}} = \frac{210,2}{25} = 8,4$$

$$\bar{R} = \frac{145}{25} = 5,8$$

LCS= 8,4+(0,577x5,8)=11,7

LCI= 8,4-(0,577x5,8)=5,1

FIGURA 17.2. Media de minutos por consulta a demanda



2.2. GRÁFICO DE PROPORCIONES (P)

Los límites de control en el gráfico p se calculan en función de n y de la desviación estándar de p. Estos límites son cambiantes si n no es constante.

El cálculo de los límites de control, se realiza según la fórmula que figura en la Tabla 17.1. Aplicándola al ejemplo de los datos que tenemos para la proporción de pacientes que tienen que esperar más de 30´ (Tabla 17.3), nos da unos valores para la desviación estándar de la proporción media de nuestro indicador =0,06, y los límites de 0,05 y 0,41. El gráfico resultante está en la Figura 17.3. Obsérvese en la fórmula que, al ser la desviación estándar de ***p*** dependiente del tamaño de la muestra (***n***), los límites de control variarían según la ***n*** de cada medición; por ello es conveniente realizar las mediciones con un tamaño de muestra constante.

TABLA 17.3. Indicador: proporción de pacientes que tienen que esperar más de 30' antes de ser atendidos. Resultados de las mediciones efectuadas en todos los días de un mes.

Muestra número	Tamaño de muestra(n)	Número de incumplimientos	Proporción de incumplimientos
1	50	12	0,24
2	50	15	0,30
3	50	8	0,16
4	50	10	0,20
5	50	4	0,08
6	50	7	0,14
7	50	16	0,32
8	50	9	0,18
9	50	14	0,28
10	50	10	0,20
11	50	5	0,10
12	50	6	0,12
13	50	17	0,34
14	50	12	0,24
15	50	22	0,44
16	50	8	0,16
17	50	10	0,20
18	50	5	0,10
19	50	13	0,26
20	50	11	0,22
21	50	20	0,40
22	50	18	0,36
TOTAL	**1,100**	**252**	

Cálculos para la línea promedio y límites de control según las fórmulas de la Tabla 17.1 para gráfico de proporciones:

$\bar{p} = \frac{252}{1,100} = 0,23$

LCS= 0,23+(3x0,06) = 0,41

LCI= 0,23- (3x0,06) = 0,05

FIGURA 17.3. Proporción de pacientes que tienen que esperar más de 30'

2.3. GRÁFICO DE VALORES INDIVIDUALES (X)

Los límites de control en el gráfico x se calculan en función del rango móvil medio ($\overline{RM}$).

Para la utilización de la distribución normal como probabilidades de referencia en lugar de la interpretación no paramétrica que veíamos en la UT 16 para los gráficos de rachas, se calcula como promedio la media de todas las mediciones (cada medición es en realidad como si fuese una muestra de un solo caso), y para los límites de control el *rango móvil medio* ($\overline{RM}$), que se calcula promediando las diferencias (en valores absolutos) entre cada medición y la inmediatamente anterior. La Tabla 17.4 ofrece un ejemplo de cálculo para el indicador de actividad del programa de cirugía menor que vimos en la UT 16. Una alternativa a la fórmula ($\overline{RM}/1,128$) como estimación de la desviación estándar es el cálculo de la propia desviación estándar para el conjunto de mediciones, como si fuese una sola muestra. Sin embargo, la estimación que resulta utilizando $\overline{RM}$ se acerca más al modelo de distribución normal en los límites que se establecen. Los límites de control que calculamos para nuestro ejemplo son 4 y 15 intervenciones respectivamente. El gráfico de control con estos límites está igualmente en la Figura 17.4.

TABLA 17.4. Cirugía menor en un centro de salud. Número de intervenciones por semana.

Semanas	Número de intervenciones	Rango móvil
1	13	—
2	7	6
3	10	3
4	8	2
5	11	3
6	12	1
7	13	1
8	12	1
9	12	0
10	8	4
11	9	1
12	5	4
13	7	2
14	9	2
15	9	0
16	5	4
17	7	2
18	8	1
19	10	2
20	11	1
TOTAL	**186**	**40**

Cálculos para la línea promedio según las fórmulas d la Tabla 17.1 para gráfico de valores

Mediana: 9

Media de intervenciones por semana:

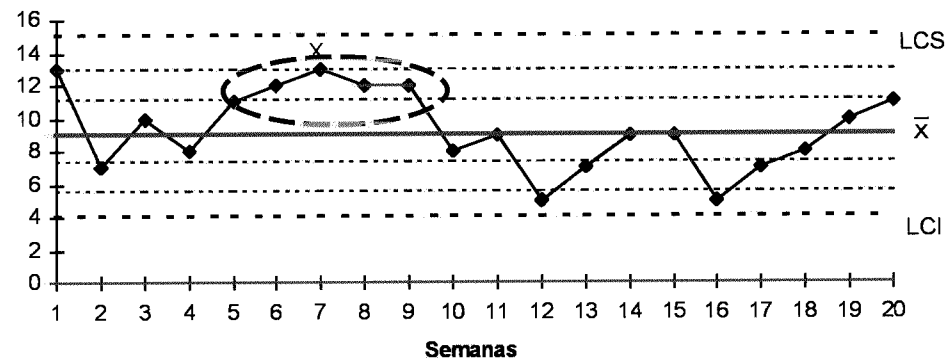
$\bar{x} = \frac{186}{20} = 9{,}3$

Rango Móvil medio: $\frac{40}{19} = 2{,}1$

$LCS = 9{,}3 + \left[3 \times \frac{2{,}1}{1{,}128}\right] = 15$

n = 20-1 = 19

FIGURA 17.4. Número de intervenciones por semana.



EMCA
Gestión de la Calidad Asistencial

2.4. GRÁFICO DE RATIOS (U)

El gráfico u utiliza la distribución de Poisson y en su plantilla los límites de control se establecen únicamente en función del promedio y el tamaño de la muestra

Los límites de control se calculan utilizando la distribución de Poisson, según la cual la desviación estándar es la propia media dividida por el tamaño de la muestra cada vez que se mide el indicador. Este es el proceder que podemos aplicar a nuestro ejemplo de número de defectos en los documentos de interconsultas (Tabla 17.5). Los límites, según la fórmula apropiada (ver Tabla 17.1) son de 0 a 1,1 defectos por documento. El gráfico completo puede verse en la Figura 17.5.

TABLA 17.5. indicador: número de defectos (criterios incumplidos) por documento de interconsulta. Resultados de 20 mediciones sucesivas en muestras de 10 documentos

Medición	Tamaño de muestra(n)	Número total de defectos	ratio
1	10	4	0,4
2	10	1	0,1
3	10	4	0,4
4	10	2	0,2
5	10	3	0,3
6	10	10	1,0
7	10	0	0,0
8	10	4	0,4
9	10	11	1,1
10	10	5	0,5
11	10	2	0,2
12	10	5	0,5
13	10	1	0,1
14	10	12	1,2
15	10	3	0,3
16	10	3	0,3
17	10	11	1,1
18	10	4	0,4
19	10	0	0,0
20	10	3	0,3
TOTAL	**200**	**88**	

Cálculos para la línea según las fórmulas de la Tabla 17.1 para gráfico de ratios:

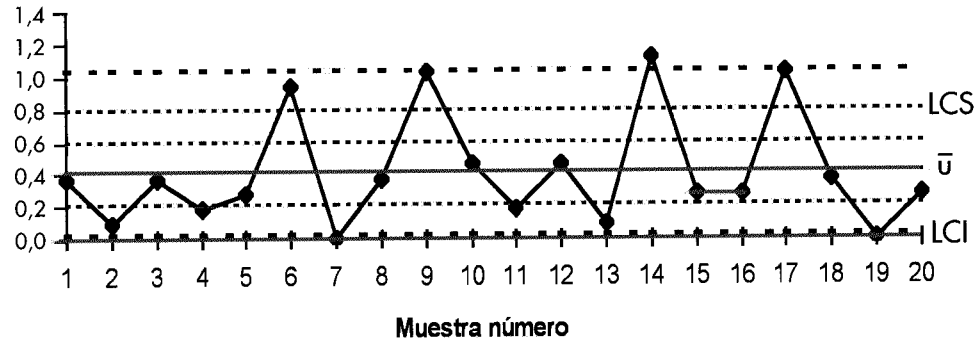
$$\bar{u} = \frac{88}{200} = 0{,}44$$

$$s = \sqrt{\frac{0{,}44}{10}} = 0{,}21$$

LCS= 0,44 + (3x0,21) = 1,1

LCI = 0,44 - (3x0,21) = 0

FIGURA17.5.Número de defectos por documento de interconsulta



3. INTERPRETACIÓN DE LOS GRÁFICOS DE CONTROL

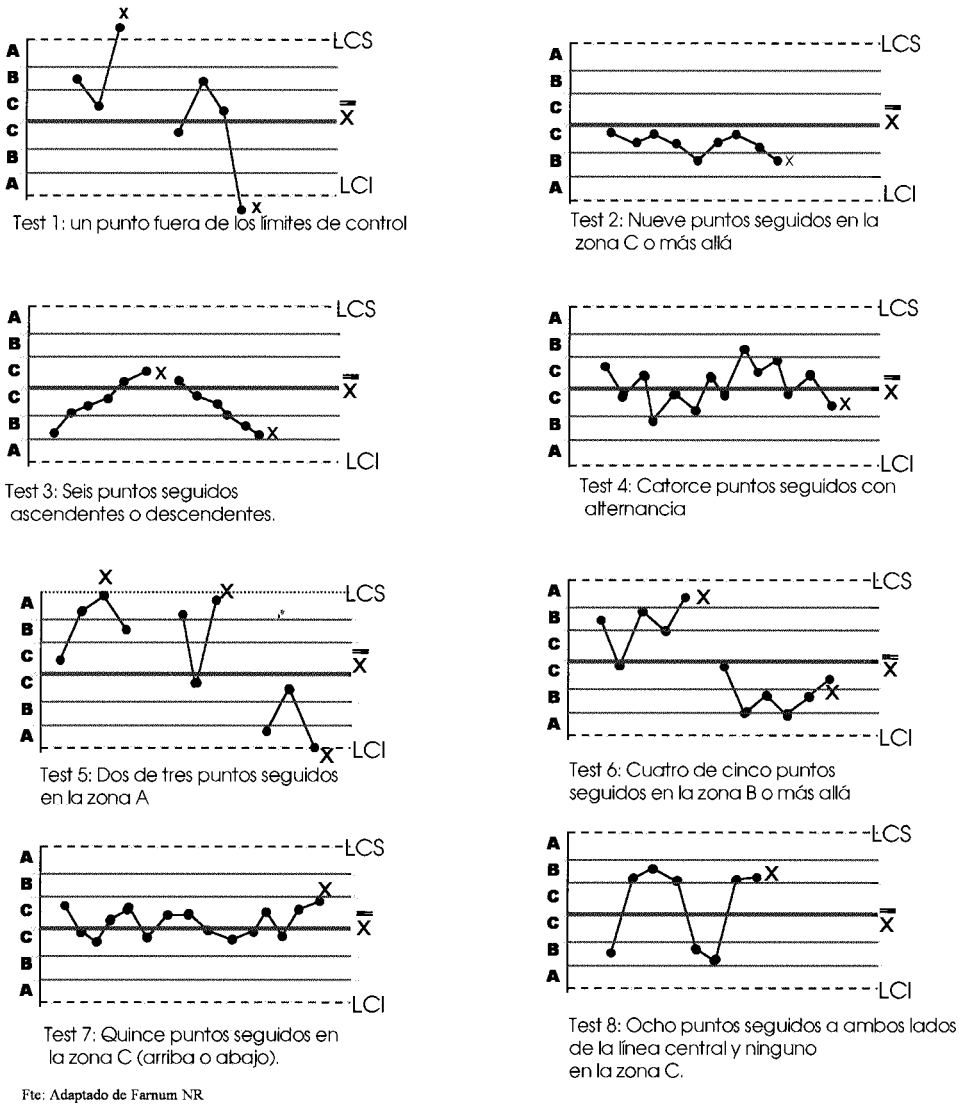
En los gráficos de control cualquier medición del indicador que caiga fuera de los límites de control es signo de problema, inestabilidad del proceso o causa "especial" que conviene investigar.

¿Qué añaden los límites de control al análisis de los datos en relación a lo que hemos visto en la UT 16 para los gráficos de rachas? En primer lugar un test adicional muy evidente: si nos resulta un punto fuera de los límites de control hemos de señalar la situación como problemática, o al menos distinta de lo esperado, significativamente "fuera de control". Revisando los gráficos de las Figuras 17.2 a 17.5 podemos ver que eso nos ocurre en dos de los indicadores: el de la proporción de pacientes que esperan más de 30' (Figura 17.3) y el del número de defectos por documento de interconsulta (Figura 17.5). Este último se nos muestra tremendamente inestable, muy lejos de estar bajo control estadístico, con varios picos en o fuera del límite de control. En el caso del tiempo de espera, habrá que averiguar qué ocurrió precisamente el día de la muestra número 15; esa causa "especial" desequilibró significativamente lo que venía siendo un comportamiento dentro de unos límites de variabilidad normales. El comportamiento irregular en los partes interconsulta debe ser también investigado (¿depende del tipo de especialista?, ¿de días?, ¿de médicos?, ¿de la demanda?....).

En los gráficos de control, son identificables una serie de secuencias significativas de "descontrol" además de los patrones de inestabilidad que se pueden apreciar en los gráficos de desarrollo

No es, sin embargo, la búsqueda de puntos fuera de los límites de control el único análisis a efectuar con los gráficos de control. En primer lugar, los mismos patrones de tendencias que vimos para los gráficos de desarrollo en la UT 16 son también aplicables a los gráficos de control. Pero además, si los "enriquecemos" señalando en la plantilla dónde se sitúan los límites correspondientes a 1 y 2 desviaciones estándar, hay otra serie de patrones que nos indicarán "descontrol", utilizando las probabilidades que corresponden a cada una de ellos. La anotación habitual para estas desviaciones estándar es hablar de "zonas" A (entre 2 y 3 desviaciones estándar), B (entre 1 y 2 desviaciones estándar) y C (entre el promedio y 1 desviación estándar). Hay establecidos al menos ocho patrones significativamente anómalos (estadísticamente poco compatibles con una situación de estabilidad), que están reproducidos en la Figura 17.6. Como cada uno de estos tests está pensado para un error α (falsos positivos) en torno a 0,3%, la *utilización conjunta* de todos ellos eleva el error a como mucho a 0,3 x 8 = 2,4%, sin embargo se incrementa dramáticamente el poder de detección de situaciones problemáticas.

FIGURA 17.6. Tests para identificar situaciones fuera de control



Test 1: un punto fuera de los límites de control

Test 2: Nueve puntos seguidos en la zona C o más allá

Test 3: Seis puntos seguidos ascendentes o descendentes.

Test 4: Catorce puntos seguidos con alternancia

Test 5: Dos de tres puntos seguidos en la zona A

Test 6: Cuatro de cinco puntos seguidos en la zona B o más allá

Test 7: Quince puntos seguidos en la zona C (arriba o abajo).

Test 8: Ocho puntos seguidos a ambos lados de la línea central y ninguno en la zona C.

Fte: Adaptado de Farnum NR

Como ilustración de esta mayor sensibilidad para detectar situaciones "fuera de control" podemos observar (Figura 17.4) como el indicador para monitorizar el programa de cirugía menor, el único de los cuatro que hemos venido analizando que parecía estar "bajo control", muestra un patrón significativamente anómalo según el test 6 de la Figura 17.6 en las semanas 5 a 9. Esta situación "fuera de control" es positiva, pero igualmente debería de ser investigada para "fijar" en el proceso aquel factor o factores que hizo que se "descontrolara" por arriba, de hecho, que mejorara su promedio.

Obsérvese que con los gráficos de desarrollo (UT 16) sólo fue posible detectar como problemático uno de los cuatro indicadores, mientras que los gráficos de control nos han permitido saber que, en realidad, ninguno de los cuatro está realmente controlado, y que ni conocemos, ni dominamos la variabilidad de los procesos que se miden. Esta situación no es la que debe de existir para una monitorización rutinaria; antes hay que conocer mejor el proceso, intervenir sobre él y llevarlo a control estadístico. Para ello eliminaremos (o incluiremos permanentemente, según sea negativo o positivo o su efecto) las causas especiales y recalcularemos los límites de control excluyendo los puntos "fuera de control" si

La división del espacio entre el promedio y los limites de control en zonas que corresponden cada una a una desviación estándar, permite a los gráficos de control tener una mayor sensibilidad para detectar situaciones problemáticas.

responden a situaciones conocidas, especiales y normales, o, tras la intervención, los cálculos para la plantilla serán sustituidos por lo que resulte de nuevas mediciones y nuevos análisis.

4. PRINCIPALES PRECAUCIONES A TENER EN CUENTA PARA EL USO DEL CONTROL ESTADÍSTICO DE LA CALIDAD

Hay tres aspectos que pueden ser con facilidad fuente de problemas al utilizar los gráficos de control. El primero de ellos, ya mencionado en la UT 16 es la confusión entre control estadístico y conformidad con los estándares, requisitos o especificaciones de calidad; el segundo es la decisión sobre el tamaño de muestra apropiado para las mediciones del indicador; y el tercero la definición del marco muestral y definición de los subgrupos (muestras) en las que se realizan las mediciones.

4.1. CONFUNDIR CONTROL ESTADÍSTICO CON CONFORMIDAD CON LOS REQUISITOS DE CALIDAD

Esta confusión, uno de los "errores que salen caros" según Deming puede revestir dos formas distintas, pero relacionadas. La primera es pensar que los límites de control son especificaciones de calidad u objetivos a cumplir. La realidad es que los límites de control se determinan estadísticamente, de forma que el gráfico de control nos dice lo que podemos esperar del proceso tal como es, no como quisiéramos que fuese. No es correcto poner directamente las especificaciones de calidad (por ejemplo el estándar y el umbral del indicador) como promedio y como límite de control.

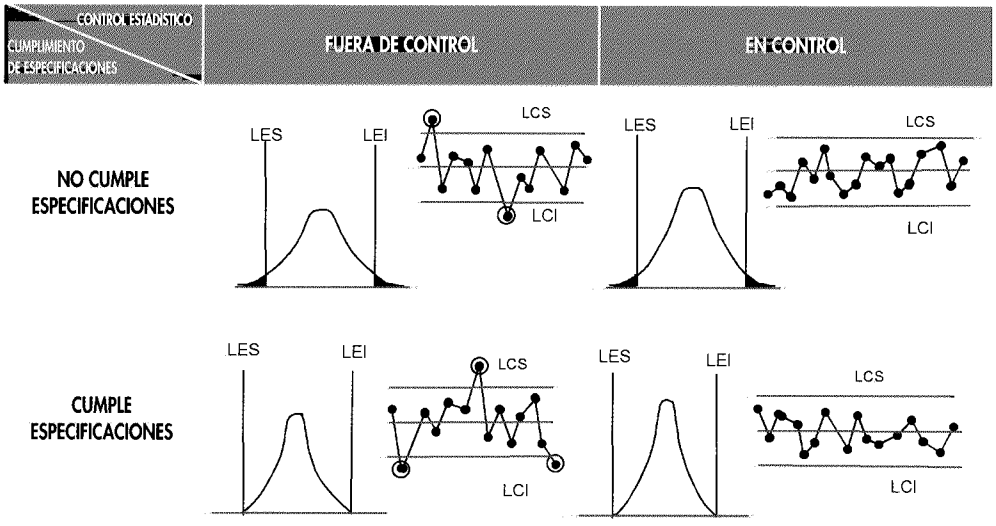
La segunda forma de caer en este error es hacer equivalente control estadístico a nivel aceptable de calidad. Hay manuales, algunos de ellos traducidos al español, que elevan este error a categoría al afirmar por ejemplo que "proceso bajo control equivale a proceso de producción capaz de ofrecer una proporción satisfactoria de elementos conformes con las especificaciones".

Si bien la inestabilidad o falta de predictibilidad estadística es siempre indeseable, la estabilidad o control estadístico no es siempre sinónimo de calidad. Un indicador puede estar estabilizado y ser estadísticamente predecible a niveles de calidad alejados del estándar deseable

La realidad es que, al igual que los gráficos de control sirven para ver si hay causa que investigar pero no qué causas, la consecución del control estadístico es más que nada un requisito para analizar el grado de conformidad con las especificaciones de calidad. Pueden producirse de hecho cuatro situaciones diferentes, tal como se refleja en la Figura 17.7 proceso estable, en control estadístico, pero que no cumple las especificaciones o estándares de calidad; proceso estable y que sí cumple las especificaciones; proceso inestable pero dentro de las especificaciones; y proceso inestable y no conforme con las especificaciones de calidad. La averiguación de la situación concreta en que nos encontramos en relación a las especificaciones de calidad puede realizarse con los gráficos de capacidad de proceso. Alternativamente, en el caso de ser un indicador cuantitativo medido como media, el análisis y control de la variabilidad interna, aun cuando el indicador se mantenga en un promedio aceptable y dentro de los límites de variabilidad esperados (límites de control), puede realizarse con los gráficos de rangos o los de desviaciones estándar (ver formulas en la Tabla 17.6) en los cuales lo que se analiza con gráfico de control es el rango o la desviación estándar de cada una de las mediciones ($\bar{s}$), el promedio o línea central del

gráfico, el rango medio de todas las mediciones ($\bar{R}$), y la media de las desviaciones estándar (Σ), recordar que promedios estables pueden enmarcar variabilidad intragrupo.

FIGURA 17.7. Control estadístico y cumplimiento de especificaciones: cuatro posibles situaciones.



Fte: Adaptado de Kume H.

Como resume Juran (citado por Deming) "los problemas importantes para mejorar comienzan una vez se ha logrado el control estadístico". Esta secuencia de: control estadístico –> conformidad con especificaciones –> monitorización, como tres actividades diferenciadas en las que pueden intervenir los gráficos de control es la que se refleja en el algoritmo (Figura 16.2) que se vió en la UT 16.

En los ejemplos de indicador que hemos venido utilizando en ésta y en la anterior UT, una vez que hayamos conseguido que los indicadores analizados sean estables y predecibles, nos deberíamos preguntar si el promedio y la variabilidad conseguidos responden a nuestros deseos en cuanto a nivel de calidad; si es así, podemos entonces utilizar el control estadístico para monitorizar y asegurarnos que nos mantenemos a ese nivel; si no es así, hay que averiguar qué cambios hay que introducir para lograr el nivel de calidad que deseamos.

La estabilidad o control estadístico es un requisito previo para plantearse el análisis de la conformidad con los requisitos o estándares de calidad de los indicadores.

4.2. TAMAÑO DE LA MUESTRA PARA LAS MEDICIONES DE LOS INDICADORES

No parece haber ninguna fórmula ni número mágico, ni siquiera un método mejor que otros para su cálculo, pero sí una serie de recomendaciones de base un tanto empírica y también, como no, probabilística.

En primer lugar hay que hacer una distinción entre variables cuantitativas y cualitativas. Para las primeras, cuya medición expresamos normalmente como media de la muestra analizada, el número de casos habitual en la industria raramente excede de 10 casos, siendo 5 casos un tamaño de muestra bastante habitual. Este número tan pequeño se compensa al ser las mediciones frecuentes y al ser la media y no valores individuales (con la correspondiente disminución

Para variables cuantitativas (gráficos) las mediciones del indicador pueden realizarse con muestras de 5 casos.

de la variación con respecto a la que podrían tener los valores individuales) lo que se utiliza para el gráfico.

En variables cualitativas (gráficos p) el tamaño de la muestra en cada medición del indicador puede ser de 20 a 40 casos, siempre que la proporción esperada de defectos o incumplimientos no sea muy pequeña (<0,05).

Para las variables cualitativas, dicotómicas, como son la mayoría de indicadores que utilizamos en la evaluación de la calidad en los servicios de salud, hay establecidas varias formas de calcular como debería ser de grande la muestra. La primera y más sencilla es pensar que es deseable que sea lo suficientemente grande como para dar oportunidad de que aparezcan los no cumplimientos o defectos en las mediciones. Fijando el número de no cumplimientos a detectar en al menos 2 en cada medición del indicador, la muestra necesaria para cada medición sería $n = 2/P_1$, donde p es la proporción esperada (promedio estable) de incumplimientos o defectos. Obviamente para proporciones esperadas mayores de 0,1 (10%) las muestras son pequeñas (para $p = 0,1 \rightarrow n = 2/0,1 = 20$), sin embargo, cuanto más pequeña sea p, mayor deberá ser el tamaño de muestra; esto es visto como una limitación importante para algunos autores, pensando, desde luego, en las bajísimas proporciones de defectos con las que se deben trabajar en la industria (para p= 1%, se necesitan muestras de 200 casos); pero no es tan problemático en los servicios de salud en los que habitualmente nos manejamos con estándares menos extremos. Por ejemplo, para una proporción media esperada de incumplimientos (equivale al estándar del indicador) de 0,05 (5%), lo cual es un valor casi extremo en nuestro medio, se necesitan muestras de 40 casos.

El tamaño de muestra necesario para los gráficos p también puede calcularse en función de la diferencia que establezcamos entre el estándar y umbral del indicador. Cuanto más pequeña sea, más casos necesitaremos.

Otra forma de calcular la muestra que necesitaríamos, es sobre la base de establecer cuál es la diferencia, en relación a la proporción media, que nos interesa identificar, con el nivel de significación que establezcamos para los tests a aplicar al analizar el gráfico. La fórmula, para una significación <0,01, una proporción de base (promedio o estándar) P_o y una diferencia a detectar "d" (por ejemplo la que estableceríamos entre el estándar y su umbral) es: $n = P_o(1-P_o)/d^2$, la cual nos da, en general, tamaños de muestra más grandes que con el otro método y, obviamente más cuanto menor sea "d".

El tamaño de la muestra en las mediciones del indicador sujeto a control estadístico es conveniente que sea constante. No obstante, los gráficos de control estadístico pueden también utilizarse, ajustándolos, en caso de que el tamaño de la muestra no sea siempre el mismo, e incluso si las mediciones son de toda la población.

Hay otros dos aspectos importantes en relación al tamaño de muestra para los gráficos de control. Uno es la conveniencia de que sea siempre el mismo, sobre todo para gráficos de proporciones y ratios. En caso contrario, los límites de control habría que recalcularlos con cada medición ajustándolos al tamaño de muestra de esa medición, según las fórmulas de la Tabla 17.1. Como alternativa, se puede utilizar para el cálculo de los límites de control un tamaño de muestra promedio, siempre que no lo estemos empleando para mediciones con muestras superiores o inferiores al 20% de este promedio.

Finalmente, aunque los grupos de control, sus fórmulas y su práctica están pensados para mediciones realizadas sobre muestras, no hay ningún inconveniente en utilizarlos para analizar y monitorizar indicadores medidos sobre poblaciones completas, siempre que se mida una evolución temporal.

El marco muestral para los indicadores a analizar con control estadístico debe ser homogéneo, para evitar que el promedio enmascare una probable variabilidad.

4.3. DEFINICIÓN DE LAS MUESTRAS (SUBGRUPOS)

Para muchos autores, la definición de los marcos muestrales y las correspondientes muestras ("subgrupos" en la terminología de la industria) en los que vamos a medir el indicador, es la parte más importante en la preparación de los gráficos de control, hasta el punto de que una definición o composición incorrecta puede conducir a gráficos inútiles.

Las mediciones deben de realizarse utilizando subgrupos "racionales", homogéneos en el sentido que exista un mínimo de variabilidad dentro del subgrupo de forma que la monitorización o el análisis pueda detectar la evolución de la variabilidad entre subgrupos (muestras). Si, por ejemplo, en un centro o unidad asistencial la actuación de los diversos profesionales es muy distinto entre sí, juntar los pacientes de todos ellos de forma indiscriminada para las muestras en las que medir el indicador puede dar una idea falsa de promedio, enmascarando la existencia de una amplia variabilidad dentro de cada muestra. La utilización de gráficos estratificados (uno para cada grupo homogéneo) desenmascara en ocasiones esta situación.

Lo que subyace en definitiva es la necesidad de ser especialmente escrupuloso con extraer muestras homogéneas en lo posible y representativas (no sesgadas). Cualquier agrupación en los marcos muestrales que implique o enmascare posibles diferencias intragrupo en métodos, personas, material o tecnología (las posibles fuentes de variabilidad) debe ser cuestionada y aclarada antes de utilizar las mediciones para análisis o monitorización con gráficos de control estadístico. Pero una vez tenidas en cuenta ésta y las demás precauciones, la incorporación de los gráficos de control estadístico a la rutina de los programas de gestión de la calidad y como forma de analizar y presentar las mediciones sistemáticas que se realizan en el sistema de salud es factible, fácil y práctica. Adicionalmente, existen varios programas de software que hacen posible el análisis rutinario de cualquier indicador, utilizando las técnicas de los gráficos de control.

La incorporación de los gráficos de control estadístico a la rutina de los programas de gestión de la calidad es factible, fácil y práctica

TABLA 17.6. Factores para los cálculos de los límites de control en gráficos de medias, desviaciones estándar y rangos

TAMAÑO DE MUESTRA	GRÁFICO PARA MEDIAS FACTORES PARA LÍMITES DE CONTROL A2	GRÁFICO PARA DESVIACIONES ESTÁNDAR FACTORES PARA LÍMITES DE CONTROL B3	B2	GRÁFICO PARA RANGOS FACTORES PARA LÍMITES DE CONTROL D3	D4
2	1,880	0	3,267	0	3,267
3	1,023	0	2,568	0	2,574
4	0,729	0	2,266	0	2,282
5	0,577	0	2,089	0	2,114
6	0,483	0,030	1,970	0	2,004
7	0,419	0,118	1,882	0,076	1,924
8	0,373	0,185	1,815	0,136	1,864
9	0,337	0,239	1,761	0,184	1,816
10	0,308	0,284	1,716	0,223	1,777
11	0,285	0,321	1,679	0,256	1,744
12	0,266	0,354	1,646	0,283	1,717
13	0,249	0,382	1,618	0,307	1,693
14	0,235	0,406	1,594	0,328	1,672
15	0,223	0,428	1,572	0,347	1,653
16	0,212	0,448	1,552	0,363	1,637
17	0,203	0,466	1,534	0,378	1,622
18	0,194	0,482	1,518	0,391	1,608
19	0,187	0,497	1,503	0,403	1,597
20	0,180	0,510	1,490	0,415	1,585
21	0,173	0,523	1,477	0,425	1,575
22	0,167	0,534	1,466	0,434	1,566
23	0,162	0,545	1,455	0,443	1,557
24	0,157	0,555	1,445	0,451	1,548
25	0,153	0,565	1,435	0,459	4,541
Fórmulas:					
LCS	$\bar{\bar{X}}+A_2\bar{R}$	$B_4 \rightarrow B_4\bar{S}$		$D_4\ \bar{R}$	
LCI	$\bar{\bar{X}}-A_2\bar{R}$	$B_3 \rightarrow B_4\bar{S}$		$D_3\ \bar{R}$	

Fte: Adaptado a partir de Walpole RE, Myers RH

BIBLIOGRAFÍA

- Peña O, Prat A. Técnicas estadísticas de control de calidad. En: ¿Cómo controlar la calidad? 2ª ed. Madrid: IMPI; 1990.
- Plsek P: Responding to variation in health care organizations. Paris: European Forum on QI in health care; 1997.
- Hanson BL, Ghare PM. Control estadístico de procesos. En: Control de calidad. Teoría y aplicaciones. Madrid: Díaz de Santos; 1990.
- Kume H. Control charts. En: Statistical methods for quality improvement. Tokyo : AOTS; 1992.
- Gitlow H, Gitlow S, Oppenheim A, Oppenheim R. Tools and methods for the improvement of quality. Boston: Irwin; 1989.
- Farnum NR. Control chart concepts. En: Modern Statistical quality control and improvement. Belmont: Duxbury Press; 1994.

18

PROGRAMAS DE MONITORIZACIÓN EXTERNA DE INDICADORES. ANÁLISIS DE PERFILES. AJUSTE Y ESTANDARIZACIÓN DE INDICADORES COMPARATIVOS

EMCA

Gestión de la Calidad Asistencial

CONTENIDO GENERAL

En esta UT se revisan los principales aspectos metodológicos en relación con la monitorización externa de indicadores. Algunos de estos aspectos (construcción/selección de indicadores, planes de monitorización, esquemas de muestreo) ya han sido tratados en otras UT, de forma que esta UT se centra fundamentalmente en lo relativo a la comparabilidad de los resultados y los métodos para las comparaciones en relación al estándar establecido y de los centros entre sí.

ÍNDICE DE CONTENIDOS

OBJETIVOS ESPECÍFICOS

1. Describir los aspectos metodológicos a tener en cuenta en la implementación de un programa de monitorización externa de indicadores.
2. Determinar la necesidad de realizar ajustes o estandarización para poder comparar los resultados.
3. Distinguir el ajuste o estandarización verdadera o "directa" del ajuste llamado "indirecto".
4. Describir los mecanismos que llevan a un ajuste directo de las tasas.
5. Comparar los resultados de un indicador con un estándar de referencia.
6. Averiguar la significación estadística de la diferencia entre el resultado de un indicador y un estándar de referencia.
7. Calcular el intervalo de confianza de la ratio estandarizada de un indicador.
8. Averiguar el grado de homogeneidad de los resultados de un indicador en un grupo de centros.
9. Identificar un estándar de excelencia relativa a un grupo de centros.
10. Analizar gráficamente los resultados de una monitorización externa de indicadores.
11. Conocer diversos métodos de análisis de la variabilidad de indicadores o tasas en pequeñas áreas geográficas.

"Para comprobar los resultados de las mediciones de un indicador en diversos centros o poblaciones, primero hay que asegurarse que son comparables."

1. INTRODUCCIÓN

En esta UT revisamos la metodología de medición y análisis comparativo de los indicadores de programas externos de monitorización de la calidad.

En esta UT nos vamos a centrar en los programas de medición y análisis comparativo de indicadores que intentan medir calidad asistencial, aunque la metodología a emplear es común a otro tipo de indicadores y programas. Esta metodología entronca con la de análisis de la variabilidad en general, un campo de investigación de primer orden, y tiene dos aspectos principales: el ajuste o estandarización de los indicadores para que puedan ser comparables, y la comparación en sí para averiguar si las diferencias entre las unidades organizativas que se comparan son relevantes y significativas. En todos los sistemas de salud y en todos sus niveles, incluso dentro de un mismo centro, podemos encontrar ejemplos de mediciones de listados más o menos extensos de indicadores; lo que no es muy común es que tanto la medición como las comparaciones se realicen con todas las precauciones y rigor metodológico que sería conveniente. En esta UT vamos a ver cómo realizar estas comparaciones de la forma más sencilla posible, a la vez que rigurosa.

2. MONITORIZACIÓN EXTERNA DE INDICADORES. JUSTIFICACIÓN Y UTILIDAD

La comparación de los resultados de un indicador tiene dos aspectos principales: la estandarización, para lograr la comparabilidad, y la comparación en sí.

En la UT 14 definimos la monitorización como la medición sistemática, repetida y planificada de indicadores de calidad, con el fin de identificar situaciones problemáticas. También vimos que, además de su validez y fiabilidad, una de las características que debe tener un buen indicador para la gestión de la calidad es su utilidad, es decir, su grado de adecuación para que puedan seguirse acciones para la mejora por parte de quienes los están midiendo, cuando se detecta un problema de calidad. Esta característica va unida al nivel y consiguiente responsabilidad para la gestión de la calidad, dentro de la estructura del sistema de salud. Cada nivel (sistema en su conjunto, áreas geográfico-administrativas en que se divida, centros asistenciales, etc.) pueden y deben manejar una serie de indicadores que les permitan controlar la calidad en relación con su responsabilidad en el sistema. El objetivo de estos programas de monitorización sería doble:

Los programas de monitorización externa de indicadores pueden servir en el nivel administrativo en que se realicen para orientar las actividades y estrategias de mejora, y en los centros monitorizados para poder comparar sus resultados con otros centros del mismo nivel

1. orientar las propias actividades y estrategias para la mejora hacia las zonas, áreas, patologías, centros o grupos de profesionales, etc. que se identifiquen como problemáticas o mejorables en relación al estándar que se establezca.
2. ayudar al siguiente (o siguientes) escalón de responsabilidad a identificar oportunidades de mejora, al comparar sus resultados con instituciones o centros asistenciales de su mismo nivel. A través de este mecanismo de provisión de datos comparativos a los programas internos de las instituciones, podrían identificarse igualmente, al menos en teoría, áreas e instituciones para proyectos de benchmarking.

La realidad, sin embargo, es muy distinta en la mayoría de los programas existentes de medición externa de indicadores. Prima sobre todo su utilización, no para gestionar la calidad, sino como mecanismo evaluador/sancionador del personal e instituciones, para reparto de incentivos, mantenimiento de contratos, o identificación en definitiva de "manzanas podridas".

La creación de Medicare en los Estados Unidos propició uno de los primeros antecedentes plenamente estructurados de programas de evaluación externa de la calidad.

Uno de los primeros esfuerzos identificables de este tipo de programas de medición externa de la calidad a nivel nacional, aunque descentralizado por regiones, vino de la mano de la creación en 1965 del programa Medicare (atención sanitaria publica para los mayores de 65 años) en Estados Unidos. Con este motivo el gobierno federal, financiador de Medicare, exigió la acreditación de las instituciones sanitarias que atendieran este tipo de pacientes y propició el establecimiento de mecanismos de control de la calidad de la atención recibida, para asegurarse que los servicios sanitarios pagados con dinero público fuesen clínicamente necesarios, apropiados, y de un nivel de calidad acorde con estándares establecidos por parte de los propios profesionales. Estos objetivos se canalizaron sucesivamente a través de los programas EMCRO (*Experimental Medicare Review Organizations*), PSRO (*Professional Standars Review Organizations*) y PRO (*Peer Review Organizations*), con unos niveles de complejidad y aspectos a evaluar sucesivamente ampliados. En el lado positivo de estas iniciativas están la creación de antecedentes en cuanto a la necesidad de monitorizar individual y comparativamente la calidad, como parte de la responsabilidad de su gestión del sistema de salud, y el desarrollo de metodologías, hoy de uso común, como es el análisis de perfiles estadísticos (que veremos más adelante). En el lado negativo o crítico está el que, en la práctica, se procuró que se orientasen más al control de la utilización de recursos y costes, que al control de la calidad en sentido amplio (e incluso estrictamente científico- técnico); y el que tuvieron un fuerte componente fiscalizador, unido incluso a la posibilidad de proponer sanciones, tanto económicas como personales.

Los programas externos de medición de la calidad puestos en marcha por Medicare tuvieron un fuerte componente de fiscalización de la utilización de recursos y control de costes, más que de la calidad en sí, pero dieron pié al desarrollo de metodologías de medición y comparación de uso común hoy en día.

Las cosas están cambiando y ya empieza a haber ejemplos de medición externa de indicadores con un objetivo más explícita y claramente enfocado a la gestión de la calidad, como las mencionadas del NHS británico y la *Veteran Health Administration* en la UT 14, o el CAHPS, e incluso algunos de los que utiliza el Plan de Calidad de Atención Especializada del INSALUD, que incluimos como ejemplo en la Tabla 18.1, o los sistemas de indicadores de otros Servicios de Salud autonómicos. En algunos casos, el programa es externo no tanto porque se mida desde fuera, sino porque la decisión de lo que hay que medir, para luego comparar, se toma desde fuera, a un nivel organizativo más alto que el de los centros asistenciales.

En los últimos años empieza a haber cada vez más ejemplos de programas de monitorización externa con el objetivo explícito de gestionar la calidad, aunque es todavía común que tengan grandes lagunas metodológicas.

TABLA 18.1. Indicadores para la monitorización de la calidad utilizados en el Plan de Calidad de Atención Especializada del Insalud.

A. CALCULADOS A PARTIR DE LOS DATOS DEL CMBD (Conjunto Mínimo Básico de Datos)
- Tasa de cesáreas.
- % de reingresos urgentes con la misma categoría diagnostica mayor, en el plazo de 30 días tras el alta.
- Mortalidad potencialmente evitable (incluye casos de asma, apendicectomía, hernia, colecistectomía, mortalidad materna y mortalidad infantil).
- Altas y estancias potencialmente ambulatorias.
- % de GDR inespecíficos e inagrupables.
- Estancia media prequirúrgica.

B. INDICADORES QUE DEBEN SER MEDIDOS EN CADA HOSPITAL.
- Tasa de reingresos en urgencias en un plazo de 72 horas,
- Tasa de prevalencia de úlceras por presión.
- Tasa de prevalencia de infección nosocomial.
- Tasa de prevalencia de infección en herida quirúrgica.
- Tasa de prevalencia de infección urinaria en pacientes sondados.
- Tasa de incidencia acumulada de infección de herida quirúrgica.
- Tasa de densidad de incidencia de infección respiratoria asociada a ventilación mecánica (UCI).
- % de estancias no adecuadas en el GDR con estancias más desviadas de la media.
- % de suspensiones quirúrgicas.
- % de pacientes que permanecen más de 3 horas en el servicio de urgencias.
- % de pacientes que permanecen más de 6 horas en el servicio de urgencias.
- % de pacientes <60 años ASA 1 a quienes se realiza radiografía de tórax en el estudio preoperatorio.
- % de pacientes > 75 años que tienen una valoración del riesgo social al ingreso en el hospital.

Elaborado a partir de: INSALUD. Plan de Calidad de Atención Especializada. Memoria 1998. Subdirección General de Coordinación Administrativa. Madrid 1999.

De cualquier forma, sea cual sea el objetivo con que se realicen las mediciones, tanto los procedimientos para la medición en sí, como la interpretación de los datos comparativos, precisan de una serie de precauciones metodológicas para que las conclusiones resulten válidas y fiables. Vamos a ver en qué consisten y cómo tener en cuenta esta serie de precauciones metodológicas, desafortunadamente no siempre controladas en los programas externos ya existentes.

3. COMPARACIÓN DE INDICADORES ENTRE CENTROS SANITARIOS. ASPECTOS METODOLÓGICOS

Sea cual sea el método empleado, la monitorización se basa en interpretaciones estadísticas de los patrones de presencia del indicador observados. En concreto se intenta realizar una interpretación de las variaciones o variabilidad que encontremos, y señalar hasta qué punto es importante y significativa. La determinación de las razones o causas que motivan esta variación eventualmente indeseable que podamos haber observado, van a tener que ser estudiadas aparte, con metodología propia de los ciclos de mejora. Lo que hacemos al comparar el mismo indicador en distintos centros es comparar patrones de comportamiento, formas de hacer las cosas, que al repetirse paciente a paciente, se puede decir

La comparación de los niveles de un indicador en distintos centros se basa en interpretaciones estadísticas de los patrones observados.

Para que la comparación de los resultados de un indicador sea válida y útil, todos los pasos anteriores, desde la construcción del indicador en adelante, deben de haberse realizado con las suficientes garantías metodológicas

que se han convertido en rutina, en un fallo (o acierto) sistemático, cuyas causas merecería la pena averiguar. Sin embargo, para llegar a estas conclusiones de una forma válida, tendremos que haber realizado la monitorización con las suficientes garantías metodológicas en cada uno de los pasos sucesivos que se resumen en la Tabla 18.2, con la peculiaridad de que un fallo evidente en cualquiera de estos pasos invalida, resta todo sentido, a los pasos sucesivos.

TABLA 18.2. Fases metodológicas para la monitorización de indicadores comparativos

FASE	PUNTOS CLAVE
1. Construcción/selección del indicador	¿Fiable?
	¿Válido?
	¿Útil para la gestión de la calidad?
2. Muestreo. Selección de casos en los que se mide	¿Aleatorio? ¿Representativo?
3. Comparabilidad de las mediciones	¿Ajustada?
	¿Factores de confusión tenidos en cuenta?
4. Selección del estándar.	¿Empírico o normativo?
	¿Razonable y realista?
5. Comparación de resultados.	¿De cada centro/área con el estándar?
	¿Entre los centros/áreas?
	¿Entre los centros/áreas en relación al estándar?
	¿Importancia de las diferencias?
	¿Significación estadística de las diferencias?

Las pautas metodológicas para las dos primeras fases (construcción/selección del indicador y selección de la muestra de casos en los que se mide) son aplicables tal y como han sido vistas en las UT 14 y 9 respectivamente, así como las características especiales al respecto, del muestreo para LQAS y control estadístico (UT 15 a 17), caso de decidirnos por estos métodos de monitorización, por lo que no vamos a repetirlas aquí. Revisaremos en cambio, como realizar o analizar críticamente, según seamos los protagonistas o receptores de este tipo de mediciones, las fases siguientes, comenzando por asegurarnos la comparabilidad de las mediciones.

4. COMPARABILIDAD DE LAS MEDICIONES

4.1. FACTORES DE CONFUSIÓN

Tal como veíamos en la UT 14 en el esquema de análisis de indicadores, la variabilidad en los resultados que se obtenga al medirlos puede depender de la manera de hacer las cosas por parte de los profesionales y/o la organización sanitaria que ofrece el servicio que evalúa el indicador, pero también de diversos factores dependientes del propio paciente, sobre los que no se tiene capacidad de influir. Estos factores del propio paciente los llamamos *factores de confusión*

para la interpretación de los resultados de la medición, porque si hay diferencias en relación a estos factores en los grupos de pacientes que atienden los centros o unidad de provisión de servicio que queramos comparar, van a producir diferencias en los indicadores, que no van a ser debidas a que el servicio valorado se realice de forma diferente, sino a las diferencias en la tipología de los pacientes atendidos. En buena lógica, si queremos comparar un indicador de calidad asistencial entre dos o más centros o grupos poblacionales, lo primero que tendremos que hacer es "depurar" los resultados de los efectos que puedan tener las diferencias en relación a los factores de confusión, para así poder comparar realmente diferencias en relación a la calidad del servicio. Los factores de confusión más frecuentes son los factores demográficos, como la edad, y el nivel de gravedad y/o co-morbilidad, factor al que se alude al hablar del case-mix de los centros y que para grupos de pacientes concretos se resume muchas veces como el grado de severidad o grado de riesgo mediante índices que resumen varios factores. Los niveles de riesgo ASA para pacientes quirúrgicos, el índice de Charlson para la co-morbilidad, la escala de Norton para las úlceras por presión, o el APACHE II (*Acute Phisiology and Chronic Health Evaluation*) para pacientes de cuidados intensivos serían ejemplos de estos índices que revisan varios factores y clasifican los pacientes según el riesgo.

Los resultados de un indicador pueden depender del nivel de calidad, pero también de factores propios del tipo de paciente, como son sus características sociodemográficas y la gravedad o riesgo de su enfermedad. Estos factores confunden la interpretación de los resultados y deben ser controlados antes de realizar comparaciones entre centros o grupos de pacientes.

La existencia de factores de confusión por los que hay que ajustar para poder comparar el mismo indicador en dos centros diferentes, es mucho más frecuente para indicadores de resultado y de consumo de recursos que para los de proceso o decisiones clínicas especificas. Por ejemplo, para comparar los niveles de satisfacción, un resultado de la asistencia, sería conveniente ajustar al menos por edad, en caso de que la composición etaria de los dos grupos a comparar fuese diferente. Igualmente es lógico pensar que habría que hacer ajustes, al menos por edad y grado de severidad, al comparar indicadores relativos a la mortalidad y frecuencia de determinadas complicaciones.

La existencia de factores de confusión es particularmente frecuente para indicadores de resultado

La primera consecuencia y recomendación que podemos expresar en base a lo que acabamos de exponer es que los factores de confusión y el tipo de ajuste a realizar es peculiar de cada indicador, y que, por tanto, hay que considerar para cada uno de ellos cuales son estos factores ligados al paciente que puedan hacer variar el riesgo de que ocurra lo que mida el indicador y, una vez identificados, ajustar o estandarizar los resultados según esos factores. ¿Cómo se hace el ajuste? Vamos a verlo a continuación, pero previamente conviene explicitar los conceptos de "factor de confusión", "riesgo", "factor de riesgo", "ajuste" y "estandarización", puesto que los estamos utilizando y los vamos utilizar profusamente en esta UT y no debe haber interpretaciones equívocas.

Cada indicador puede tener sus propios factores de confusión, que hay que tratar de identificar y controlar para cada situación en concreto.

Un *factor de confusión* es una característica o circunstancia sobre la que no se puede intervenir (por ejemplo, edad o sexo), o sin interés para el estudio que queremos realizar, pero que se asocia tanto al individuo en el que realizamos la medición como a la variable o efecto que se mide (en nuestro caso el nivel de calidad), de forma que la variabilidad en la presencia de los factores de confusión hace variar también los resultados. Por eso, para comparar el efecto que nos interesa (por ejemplo nivel de calidad) entre los grupos de pacientes, nos conviene controlar o eliminar el efecto producido por el factor (o factores) de confusión. Esto es lo que se llama "ajuste" o "estandarización".

Un *factor de riesgo* es un término epidemiológico referido a una característica o circunstancia, normalmente modificable, que se asocia al individuo en que

Para medir calidad, los factores de riesgo en sentido epidemiológico, se consideran factores de confusión.

medimos y a la variable o efecto de interés (normalmente la aparición de una enfermedad o característica indeseable), influyendo en el "riesgo" de que aparezca. La palabra "riesgo" es equivalente a probabilidad de que ocurra el efecto en estudio. Los estudios epidemiológicos se centran con frecuencia en la identificación de los factores de riesgo para una determinada patología o efecto de morbi-mortalidad que se haya definido como objeto de estudio.

Sin embargo, al medir y comparar indicadores de calidad, lo que en estudios epidemiológicos serian "factores de riesgo", por ejemplo para mortalidad o aparición de complicaciones, pueden ser considerados "factores de confusión", porque influyen en el resultado de nuestro interés confundiendo el efecto de mayor o menor calidad del servicio ofrecido, que es el que nos interesa. Por tanto, va a ser frecuentemente necesario ajustar los resultados de la medición de indicadores de calidad tanto por los factores de confusión propiamente dichos como por los factores de riesgo.

4.2. AJUSTE O ESTANDARIZACIÓN DE INDICADORES

Ajustar o estandarizar significa eliminar el efecto de los factores de confusión, para poder comparar los resultados de un indicador en dos o más grupos de pacientes.

Ya hemos indicado que "ajustar" significa controlar o eliminar el efecto de los factores de confusión para poder comparar adecuadamente las mediciones del indicador en dos o más grupos de pacientes (dos o más centros, áreas, etc.), en términos de niveles de calidad atribuibles a la forma de actuar de los servicios de salud. Al "ajuste" se le llama también "estandarización" porque lo que se hace es calcular el valor que tendrían los resultados del indicador a comparar en una situación "estándar", con una distribución conocida de los factores de confusión por los que se quiere ajustar. Los valores "estandarizados" que se estimen para cada centro o grupo de pacientes van a ser entonces comparables, al haber eliminado las diferencias debidas a las diferencias a su vez en los factores de confusión por los que se ha realizado el ajuste.

En algunos manuales se sigue distinguiendo entre estandarización "directa" e "indirecta" como dos procedimientos igualmente válidos. Sin embargo la llamada estandarización "indirecta" sólo sirve para comparar cada población con el estándar de referencia pero no con otras poblaciones.

Este proceder de aplicar los valores de los indicadores a comparar a una población estándar es lo que se conoce en algunos manuales como ajuste o estandarización "directa", distinguiéndola de la llamada estandarización "indirecta", en la cual se procede a la inversa: lo que se hace es aplicar los valores que tiene el indicador en una población que se considera estándar, a los diversos grupos o poblaciones a comparar. El que se siga hablando de ajuste o estandarización "indirecta" en los manuales y, más aún, que sigan publicándose artículos incluso en revistas de prestigio, que la hayan utilizado, es una curiosidad científico/sociológica muy difícil de explicar, dado que la estandarización indirecta no sirve para comparar los resultados de dos o más poblaciones eliminando el efecto de los factores de confusión para los que se pretende ajustar, porque es cada una de las poblaciones la que se utiliza de estándar en cuanto al factor de confusión para el cálculo de "su" tasa estandarizada, de forma que sólo sería comparable en todo caso con la población de la que se toman las tasas "estándar". Vamos a ver con un ejemplo simple como funcionan ambos procedimientos.

a) Ajuste "directo" de los resultados de un indicador

Supongamos que queremos comparar la tasa de infección nosocomial de dos

hospitales. Como sabemos aparte de la calidad con que se realice la atención hospitalaria, hay muchos factores que pueden influir por sí solos en la tasa de infección (factores de confusión), que sería conveniente controlar para que la comparación pueda referirse efectivamente a diversos niveles de calidad. Supongamos, sin embargo, que vamos a ajustar inicialmente por sólo uno de estos factores: la proporción de pacientes médicos, quirúrgicos, de ginecología-obstetricia y de cuidados intensivos que hay en cada uno de los dos hospitales, dado que el riesgo de infectarse, incluso haciendo bien las cosas, es diferente en función de este factor. Los datos de ambos hospitales aparecen en la Tabla 18.3.

TABLA 18.3. Tasas de infección según Hospital y Servicio Clínico

	HOSPITAL A			HOSPITAL B		
	PACIENTES	INFECCIONES	TASA (%)	PACIENTES	INFECCIONES	TASA (%)
Medicina	350	24	6,9	400	28	7,0
Cirugía	450	36	8,0	350	28	8,0
Gine-Obstetricia	100	4	4,0	200	8	4,0
UCI	100	25	25,0	50	13	26,0
Total	1.000	89	8,9	1.000	77	7,7

En estos datos puede verse que la tasa total sin ajustar (también llamada "bruta" o "cruda") del hospital A (8,9%) es mayor que la del hospital B (7,7%); sin embargo las tasas específicas por tipo de paciente (el factor de confusión para el que queremos ajustar) son prácticamente idénticas, sólo hay una ligera diferencia en los pacientes de medicina y de UCI, que son mayores en el hospital B, así que las diferencias que observamos en la tasa total sin ajustar se debe precisamente a la diferente proporción de cada tipo de paciente que existe en cada hospital. Para eliminar este factor de confusión en la comparación, vamos a calcular cual sería la tasa resultante para cada hospital si la proporción de cada tipo de pacientes fuese idéntica en los dos. Para ello elegimos una distribución que utilizaremos como estándar para los dos hospitales, a la cual le vamos a aplicar las tasas específicas de cada tipo de paciente en cada hospital. La población o distribución estándar será la que figura en la siguiente Tabla 18.4, y el número de infecciones en cada tipo de paciente, aplicando las tasas específicas de cada hospital serían:

TABLA 18.4. Tasas de infección ajustadas por población estandar

	POBLACIÓN ESTÁNDAR	INFECCIONES SEGÚN LAS TASAS DEL HOSPITAL A	INFECCIONES SEGÚN LAS TASAS DEL HOSPITAL B
Medicina	4.000	276	280
Cirugía	4.000	320	320
Gine-Obstetricia	1.000	40	40
UCI	1.000	250	260
TOTAL	10.000	886	900
Tasa Total ajustada (%)		8,86	9,0

En el llamado ajuste "directo" se aplican las tasas específicas de cada estrato a ajustar, a una población, llamada estándar. La tasa global que resulta en la población estándar al realizar esta operación para cada una de las poblaciones a comparar, es la tasa "ajustada", y comparable.

Lo que hemos hecho es, para cada tipo de paciente de la población estándar, calcular el número de infecciones que habría si tuviese la tasa específica para ese estrato, de cada hospital. Así por ejemplo, para los pacientes de Medicina, el hospital A tiene un 6,9% de infecciones, lo cual nos daría 276 infecciones en 4.000 pacientes (276 es el 6,9% de 4.000). Cuando se han hecho estos cálculos para todos los tipos de paciente en ambos hospitales, sumamos el número total de infecciones resultante y calculamos la tasa general, para cada hospital, esta vez ya ajustada por tipo de paciente, utilizando como numerador el número total de infecciones que hemos calculado en la población estándar para cada hospital (886 y 900 respectivamente), y como denominador el total de la población estándar (10.000 pacientes).

Como puede verse, el ajuste (eliminar las diferencias en el factor de confusión entre los centros a comparar) nos ha revelado la situación real: las dos tasas son prácticamente iguales, si acaso con una ligera diferencia (tasa mayor en el hospital B) que refleja a su vez la ligera diferencia que hay en las tasas de infección de los pacientes de medicina y UCI en el hospital B.

El procedimiento que hemos seguido es el que se conoce como estandarización o ajuste directo, cuyos pasos en general son los siguientes:

1° Estratificar los datos de los centros o poblaciones a comparar en función del factor de confusión a controlar (en nuestro ejemplo, este factor ha sido el tipo de paciente), calculando la tasa específica en cada estrato.

2° Elegir la población estándar, estratificada según el factor de confusión a controlar.

3° Calcular el número de casos que resultarían en la población estándar, aplicando las tasas específicas por estrato de los centros a comparar.

4° Calcular las tasas (ajustadas) sumando el número de casos que corresponde a cada centro en la población estándar, y dividiéndolo por el total de población de la población estándar.

La población estándar suele ser una población conocida (región, nación, etc.) con la que nos puede interesar también la comparación

Habitualmente se elige como población estándar una población conocida que pueda servir a su vez de referencia para comparar, como puede ser la del área, región o de todo el país. Pero el procedimiento es idéntico.

Los ajustes por más de dos o tres factores se realizan con modelos multivariables.

Cuando hay que ajustar por más de un factor, pueden establecerse tantas categorías específicas como resulten de la combinación de factores, o ajustarlos sucesivamente (por ejemplo las tasas a ajustar por tipo de paciente en nuestro ejemplo, pueden estar ya ajustadas por edad). Sin embargo cuando son varios los factores, se construyen modelos multivariables que los incluyen todos. Los índices de severidad serian en cierto modo un ejemplo de ello.

b) Ajuste "indirecto" de los resultados de un indicador

¿Y cómo es el ajuste o estandarización llamada "indirecta"? Pues consiste en tomar de la población estándar, no la distribución en cuanto al factor de confusión a eliminar, sino las tasas específicas por cada estrato de este factor, que luego se aplican a cada uno de los centros o poblaciones a comparar y se calcula con ello la tasa resultante, que estaría presumiblemente "ajustada". Este tipo de ajuste hay quien lo recomienda en los casos en que las tasas específicas por los estratos del factor a comparar son "inestables", es decir cuando el numerador

o el denominador son pequeños, y pequeños cambios repercutirían en una variación desproporcionada en la tasa, o bien cuando no se conocen en las poblaciones a comparar las tasas específicas por cada estrato del factor por el que se pretende ajustar. Como cálculo adicional, en un intento probablemente de hacer más plausible la comparabilidad de las tasas "ajustadas" resultantes, se calcula la "ratio ajustada" dividiendo la tasa sin ajustar por la "ajustada" que hemos calculado en cada centro a comparar. Con esta nueva "estandarización" (que en el caso de tasas de mortalidad se conoce como "ratio de mortalidad estandarizada" (*Standarized Mortality Ratio, SMR*), se compara cada población con el estándar (lo cual es posible, porque hemos tomado para ambos casos como distribución común del factor de confusión a eliminar, el propio de la población en estudio); pero también se pretende comparar con la otra población (o poblaciones) en las que hayamos procedido a realizar este tipo de ajuste, lo cual no tiene ningún sentido porque en cada caso se ha utilizado como estándar, en relación a la distribución del factor de confusión, la propia de cada población a comparar. Este procedimiento no tiene en cuenta este importante hecho: las diferencias en el factor de confusión entre las poblaciones a comparar siguen estando presentes en la tasa pretendidamente "ajustada", ni tampoco que las tasas ajustadas que se calculan para comparar son siempre tasas ficticias cuyos valores tiene sentido relativo, sólo para comparar, pero que dependen de la distribución específica del factor de confusión en la población elegida como estándar (que es única en el caso de la estandarización directa, pero que es múltiple, cada población hace de estándar de sí misma, en la "indirecta"). Vamos a verlo de forma empírica, siguiendo con nuestro ejemplo.

En el llamado ajuste "indirecto" lo que se toma de la población estándar son sus tasas específicas, dejando intactas en cada población a comparar, su propia distribución en relación al factor de confusión.

Supongamos que queremos hacer la misma comparación que hemos realizado con ajuste directo: tasas de infección hospitalaria en dos hospitales, ajustando por el tipo de pacientes. Los datos de los hospitales a comparar son los mismos a los utilizados en la Tabla 18.3, pero el estándar va a ser esta vez unas determinadas tasas especificas por cada estrato del factor a controlar, en vez de una distribución de la población según este factor. Supongamos que en el estándar que vamos a utilizar es una tasa específica del 10% para todos los estratos menos UCI, que será 15%. Para el ajuste "indirecto" aplicamos estas tasas del estándar a los dos centros a comparar para obtener los casos "esperados". Después dividiremos los casos "observados" originalmente (sin ajustar), por los "esperados", obtenidos al aplicar las tasas específicas del estándar, lo cual es lo que se llama su "ratio estandarizada". Los resultados en nuestro ejemplo aparecen en la Tabla 18.5.

TABLA 18.5. Casos observados y casos esperados en 2 hospitales

	HOSPITAL A		HOSPITAL B	
	Casos observados	Casos esperados	Casos observados	Casos esperados
Medicina	24	35	28	40
Cirugía	36	45	28	35
Gine-Obstetricia	4	10	8	20
UCI	25	15	13	7,5
TOTAL	89	105	77	102,5
Ratio estandarizda	$\frac{89}{105} = 0{,}85$		$\frac{77}{102{,}2} = 0{,}75$	

La columna de los casos esperados se ha obtenido aplicando las tasas del estándar a la población de los hospitales a comparar. Así por ejemplo para los pacientes de Medicina, 35 es el 10% de 350 (Hospital A) y 40 es el 10% de 400 (Hospital B).

Como puede verse, la conclusión a la que llegamos al comparar por este método las tasas de infección de los hospitales A y B no tiene nada que ver con la realidad. Lo único que podríamos concluir (a falta de comprobar la significación estadística de esta afirmación) es que las tasas de ambos hospitales son inferiores a las del estándar (si fuesen iguales la ratio ajustada sería 1), pero la comparación entre ellos no es posible porque hemos mantenido en cada hospital su propia distribución en cuanto al factor de confusión. Además las tasas ajustadas son siempre tasas ficticias, calculadas para comparar, y su magnitud depende de la distribución en la población estándar del factor de confusión por el que se ajusta, por lo cual sólo serían comparables tasas ajustadas con arreglo al mismo estándar en cuanto al factor de confusión. Si no se ha utilizado para todos los centros a comparar el mismo estándar en cuanto a la distribución del factor de confusión, como es el caso del ajuste "indirecto" que utiliza la distribución de cada centro tanto para los casos "observados" como para los "esperados", la comparación no tiene ninguna lógica. (nota: una discusión más detallada de esta cuestión puede verse en el capítulo 5 de Rothman KJ. Modern Epidemiology. Boston: Little, Brown and Co.; 1986).

La comparación entre las tasas de varias poblaciones ajustados por el método indirecto no tiene ninguna lógica.

Conviene retener, sin embargo, el procedimiento de comparar casos observados a casos esperados según el estándar de referencia, porque es probablemente la forma más sencilla e intuitiva para ver diferencias entre un centro y el estándar de calidad que se establezca: si no hay diferencia la ratio sería 1; si es mayor en el centro estimado que en el estándar, la ratio será >1, y si es menor en el centro en relación al estándar, la ratio será <1. Esta misma comparación en relación al estándar de calidad puede hacerse con varios centros o grupos poblacionales, tal como veremos más adelante, siempre que las tasas a comparar hayan sido previamente ajustadas.

Nos hemos detenido en los detalles del ajuste de tasas, ilustrándolo con un ejemplo simple, con la intención de sustanciar lo más clara y convincentemente posible las siguientes recomendaciones, tanto si somos nosotros quienes hacemos la monitorización como si somos "clientes" de la misma:

La llamada "ratio estandarizada" es útil para comparar cada población con el estándar de referencia.

1. Las comparaciones entre centros deben realizarse en base a valores ajustados por los factores de confusión que sepamos influyen en lo que se mide.
2. El ajuste debe realizarse por el método conocido como "directo", es decir utilizando un mismo estándar de referencia en cuanto a los factores de confusión, para todos los centros a comparar.
3. Para poder ajustar correctamente, ha que realizar concurrentemente la medición del indicador a comparar y de los factores de confusión por los que queremos ajustar. De ahí la importancia de reflexionar sobre ellos e identificarlos de antemano a la hora de diseñar o seleccionar el indicador, tal como vimos en la UT 14.

La comparación de indicadores sin ajustar, si el ajuste se ve necesario, o ajustados por el método "indirecto" son de una credibilidad perfectamente cuestionable.

4. La comparación de indicadores sin ajustar entre centros, áreas, regiones, o grupos de pacientes en general, a menos que se justifique la inexistencia de factores de confusión, o ajustados por el método "indirecto", son de una credibilidad perfectamente cuestionable.

5. SELECCIÓN DEL ESTÁNDAR DE CALIDAD

El siguiente componente necesario para la monitorización va a ser determinar el estándar de calidad, nivel de cumplimiento del indicador, que vamos a tomar como referencia para averiguar si los resultados en los distintos centros, áreas o poblaciones es o no problemática.

Es muy frecuente en los programas externos la utilización de estándares empíricos, normalmente la media del área geográfica en que se encuadran los centros a monitorizar. Es también posible establecer el estándar de forma normativa, en base a revisiones bibliográficas o en función de la importancia que tenga el que se alcancen determinados niveles (piénsense por ejemplo en estándares de mortalidad, complicaciones, cobertura vacunal, etc.).

El estándar de referencia en los programas de monitorización externa suele ser el promedio empírico del conjunto de centros monitorizados, pero también puede establecerse un estándar normativo, o buscar un estándar de excelencia relativa (distinto del promedio) en el grupo de centros monitorizados.

Una forma particular de establecer un estándar empírico son los estándares de excelencia relativa identificados entre los propios centros que se comparan, cuya variante extrema serían los valores de nivel más alto que se encuentren, resultando sin embargo más razonable buscar un estándar más homogéneo, aunque igualmente deseable, con los procedimientos que veremos más adelante. Esta alternativa de "excelencia interna" no es la media de los centros que se monitorizan, pero sigue siendo realista, que es una de las características que deben tener los estándares que se elijan. Si los estándares no son realistas y razonables, el considerarlos como punto de referencia para valorar la calidad no va a parecer lógico en los centros a monitorizar, y puede no contribuir adecuadamente a motivar para la mejora.

Una vez elegido el estándar y tenidos en cuenta los posibles factores de confusión en la medición y cuantificación del indicador, estamos en condiciones de comparar los resultados.

Sea cual sea su origen, el estándar que se elija debe ser realista y razonable.

6. COMPARACIÓN DE RESULTADOS DE MONITORIZACIONES EXTERNAS

Ante los resultados de una monitorización en varios centros, áreas o unidades de provisión de servicios de que se trate, podemos estar interesados en realizar diversos tipos de comparaciones. Puede interesarnos, por ejemplo comparar cada uno de los centros con el estándar (¿cuáles cumplen y cuáles no?, ¿cuáles están significativamente mejor y cuáles significativamente peor?), o bien los centros entre sí (¿hay diferencias significativas entre ellos?, ¿cómo se agrupan en relación al estándar?). Pero en todos los casos nos va a interesar calibrar la importancia de las diferencias y también muy probablemente saber el grado de significación de las mismas. Para todo ello se han desarrollado, y se siguen desarrollando, diversas metodologías, en relación sobre todo a la comparación de la variabilidad de indicadores poblacionales en pequeñas áreas geográficas (small area analysis). Aunque haremos referencia a todos ellos, vamos a detallar los más sencillos y de utilidad más común para comparar los datos de un indicador con un estándar de referencia. Nos vamos a referir al caso en que hemos realizado estimaciones del valor del indicador, que aún es la situación más frecuente. No consideraremos las metodologías del LQAS y del Control Estadístico de la Calidad, útiles ambas para monitorización pero que no se basan en estimaciones del nivel de cumplimiento.

Son varios los tipos de comparación que podemos estar interesados en realizar, pero en todos los casos vamos a calibrar la importancia de las diferencias y su grado de significación estadística.

Nos vamos a centrar en la comparación de estimaciones del nivel de cumplimiento de los indicadores.

6.1. COMPARACIONES DE LOS INDICADORES DE CADA CENTRO O UNIDAD DE SERVICIO CON UN ESTÁNDAR DE REFERENCIA

La comparación entre los indicadores de cada centro y el estándar de referencia puede realizarse con varios métodos, muchos de ellos sobre la base de la relación entre los valores observados y los esperados si cada centro tiene el nivel del estándar.

Las comparaciones con el estándar las vamos a realizar viendo la relación de los valores observados (previamente ajustados, cuando queramos comparar más de un centro o grupo poblacional) con los que serían de esperar en el centro en estudio si se diesen los valores del estándar de referencia, a los que llamaremos valores "esperados". La base de la comparación puede ser la diferencia entre los valores observados y los esperados, o bien la ratio entre ellos. También pueden realizarse comparaciones más elementales, como el llamado análisis de perfiles estadísticos, encaminadas fundamentalmente a detectar valores extremos. Vamos a ver cada uno de estos tres métodos, comenzando por este último.

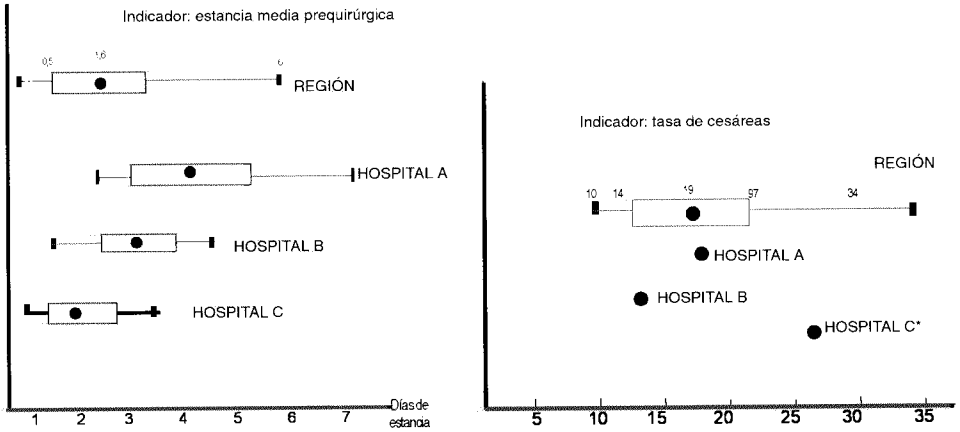
a) Análisis de perfiles estadísticos

El análisis de perfiles estadísticos es la manera más simple de comparar y suele realizarse de forma gráfica. Tiene como principal objetivo la identificación de valores extremos.

El análisis de perfiles estadísticos es un análisis habitualmente gráfico que suele tomar como valores de referencia los correspondientes al conjunto regional nacional o del conjunto de los centros analizados, y que se refleja en un gráfico por ejemplo tipo "box and whisker" (explicado en la UT 11), en el que se comparan los valores y gráficos correspondiente de cada uno de los centros. Se pueden considerar problemáticos y dignos de análisis más detallado, aquellos centros con valores por encima del percentil 75 ó por debajo del percentil 25 (según sea el signo del indicador), y tanto más cuanto más alejado se encuentre. La Figura 18.1 muestra dos ejemplos de análisis de perfiles estadísticos, uno para un indicador tipo media (estancia media preoperatoria) y otro tipo tasa (tasa de cesáreas).

Como ya hemos dicho, este tipo de análisis tiene como principal objetivo identificar valores extremos, y, secundariamente visualizar la variabilidad existente. Suele tener más interés, sobre todo para los programas internos y para ámbitos geográfico-administrativos regionales o infraregionales una comparación detallada de cada uno de los centros con el estándar de referencia.

FIGURA 18.1. Análisis de perfiles estadísticos mediante gráficos tipo box and whiskers



** El hospital A presenta valores medios por encima del percentil 75 de la Región y debe ser analizado en detalle.*

** El hospital C presenta un promedio por encima del percentil 75 y debe ser analizado en detalle.*

b) Comparación de la diferencia entre el valor observado y el estándar de referencia

Los valores "observados" se refieren al número de eventos que mide al indicador (su numerador), que calcularemos de forma ajustada si queremos comparar más de un centro. Los valores "esperados", con los que vamos a comparar los observados, se calculan aplicando al centro a comparar las tasas que consideramos estándar. Vemos entonces la diferencia entre valores observados y esperados, y podemos saber de una forma muy simple si esta diferencia es significativa, calculando la desviación estándar que corresponde a la tasa esperada, y viendo cuantas desviaciones estándar representa la diferencia; el resultado de esta operación es un valor de z, de forma que si es >1,96 la diferencia es significativa con una probabilidad de equivocarnos inferior al 5%, y si es >3, la diferencia es significativa con una p<0,001. Puede también verse el valor exacto de p (significación estadística o probabilidad de equivocarse) buscando en una tabla de la distribución normal (dos colas) el que corresponde a la z que nos salga en los cálculos.

Para comparar un valor observado en un centro con el estándar de referencia, se puede calcular la diferencia entre los casos observados y los esperados si el centro tuviese las tasas del estándar. Dividendo esta diferencia por la desviación estándar de la tasa esperada podemos saber si es significativa.

La Tabla 18.6 contiene el resultado de todos los cálculos necesarios para la comparación de la tasa de cesáreas en cinco hospitales, en relación con la tasa promedio nacional que se toma como estándar. Los datos no son reales, pero sirven como ilustración.

TABLA 18.6. Tasas de cesáreas en cinco hospitales

	n	CESÁREAS OBSERVADAS (O)	TASA OBSERVADA O/n	CESÁREAS ESPERADAS (E)	TASA ESPERADA E/n	DE	O-E	O/E	Z
H_1	288	44	15,3	68,9	23,9	7,2	-24,9	0,64	-3,44
H_2	585	150	25,6	138,3	23,6	10,3	11,7	1,08	1,14
H_3	141	35	24,8	36,7	26,0	2,5	-1,7	0,95	-0,33
H_4	269	38	14,1	56,4	21,0	6,7	-18,4	0,67	-2,76
H_5	225	35	15,6	50,9	22,6	6,3	-15,9	0,69	-2,53

* Los valores observados están ajustados por el riesgo en los cinco hospitales. Los valores esperados corresponden a los que tendría cada hospital si tuviese las tasas específicas medias a nivel nacional, que se utilizan como estándar.

El procedimiento paso a paso es como sigue:

1. Calcular el número de casos de cesárea observados (O) ajustados por riesgo, y el número de cesáreas esperados (E), aplicando a cada hospital las tasas de cesáreas del estándar por un procedimiento semejante al que hemos visto para el ajuste llamado "indirecto".
2. Calcular la tasa observada (O/n) y la tasa esperada (E/n).
3. Calcular la desviación estándar de la tasa esperada $S_E = \sqrt{n.p.(1-p)}$ donde n es el número de partos (denominador) de cada hospital y p es = E/n, es decir la tasa de cesáreas esperada, expresada como proporción.

5. Calcular la z que corresponde, dividiendo la diferencia entre casos observados y esperados, por la desviación estándar

$$z = \frac{O - E}{\sqrt{n.\,p(1 - p)}}$$

El nivel de significación estadística que establezcamos como aceptable para cada comparación, es conveniente que se calcule teniendo en cuenta el número de comparaciones que vayamos a hacer

6. Concluir sobre la significación estadística de la diferencia en función del valor de z. En general para $p<0,05$, z ha de ser $\geq 1,96$.

En relación al grado de significación estadística, diversos autores señalan que si vamos a realizar múltiples comparaciones (una para cada centro valorado), es conveniente establecer un nivel de significación más alto, dado que cuanto más comparaciones hagamos simultáneamente, más probabilidad hay de que obtengamos algún valor significativo debido al azar. En concreto, la probabilidad de encontrar al menos un valor significativo al azar, según el número de comparaciones que realicemos (C) y nivel de significación que escojamos para cada una (α) es: $\alpha.C$

De forma que, por ejemplo, si realizamos 5 comparaciones y vemos nivel de significación $\leq 0,05$ en cada una, la probabilidad de encontrar por azar al menos una de ellas significativa es: (5). (0,05)= 0,25 ó 25%.

Si quisiéramos asegurarnos con una significación de $\leq 0,05$, que todas las comparaciones significativas realmente lo son, deberíamos exigir un nivel de significación en cada comparación, que llamaremos α; tal que, en nuestro ejemplo de 5 comparaciones: α sea $\leq 0,05$, lo cual nos da.

$$\alpha \leq \frac{0,05}{C} \leq 0,01$$

En general, para asegurarnos una significación $\leq 0,05, \alpha \leq \frac{0,05}{C}$, donde α' será el nivel de significación que vamos a exigir a cada comparación y C el número de comparaciones que hagamos; con ello tendremos una probabilidad máxima del 5% de que alguna de las comparaciones aparezca significativa sólo por azar. La Tabla 18.7 resume los cálculos realizados para el nivel de significación a exigir en cada comparación, y los correspondientes valores de z, según el número de comparaciones que realicemos.

Una precaución adicional a tener en cuenta es que este procedimiento tiene validez estadística siempre que todos los valores esperados sean ≥ 5.

TABLA 18.7. Nivel de significación estadistica aconsejado para cada comparación, según el número de comparaciones que se realicen

Nº COMPARACIONES	SIGNIFICACIÓN PARA CADA UNA	VALOR DE Z	SIGNIFICACIÓN GENERAL ASEGURADA
1	$\leq 0,05$	$\geq 1,96$	$\leq 0,05$
2—5	$\leq 0,01$	$\geq 2,58$	$\leq 0,049$
6—50	$\leq 0,001$	$\geq 3,27$	$\leq 0,049$
k	$\alpha' = \frac{\alpha}{k}$	Ver tabla de distribución normal	$\leq \alpha$

Vamos a ver como realizamos todos estos pasos para la comparación con el estándar en uno de los hospitales de la Tabla 18.6. Tomemos por ejemplo el hospital 1 (H_1), con 44 cesáreas observadas (O) en 288 partos (n), mientras que por el estándar nacional serían de esperar, con el mismo nivel de riesgo, 68,9 (E). La diferencia (O-E) es −24,9, y la desviación estándar de 23,9%, que sería la tasa esperada, es:

$$\sqrt{n\,(p)\,(1-p)} = \sqrt{288\,.\,(0{,}239)\,.\,(0{,}761)} = 7{,}2$$

y el valor de

$$z = \frac{O-E}{S_E} = \frac{-24{,}9}{7{,}2} = -3{,}44$$

Este valor es significativo, es decir en el hospital 1 se realizan menos cesáreas, ajustadas por riesgo, que en la media nacional, tanto si tenemos en cuenta que es una de cinco comparaciones (necesitaríamos una $z \geq 2{,}58$, ver Tabla 18.7), como si fuese la única comparación. Similares cálculos se han hecho para los hospitales 2, 3, 4 y 5, resultando significativamente menores que la tasa nacional H_1 y H_4.

c) Comparación en base a la ratio estandarizada.

Un procedimiento alternativo de comparación con el estándar es calcular el intervalo de confianza de la ratio estandarizada (O/E). El razonamiento es simple y muy fácil de interpretar:

- Si el intervalo contiene el valor 1, no hay diferencia entre el valor en estudio y el estándar, ambas tasas observada y esperada son iguales o, al menos, no significativamente diferentes.
- Si el límite inferior del intervalo es >1, la tasa en estudio es significativamente mayor que el valor estándar.
- Si el límite superior del intervalo es <1, la tasa en estudio es significativamente menor que el valor estándar.

Un procedimiento alternativo para comparar es calcular el intervalo de confianza de la ratio estandarizada.

¿Cómo se calcula el intervalo de confianza? La fórmula la propusieron Breslow y Day, y ha sido posteriormente refinada por Lawthers, Palmer, Edwards et al. en la Escuela de Salud Pública de Harvard, introduciendo un método jerárquico que fue utilizado en el proyecto DEMPAQ (Develop and Evaluate Methods for Promoting Ambulatory Care Quality).

La fórmula inicial para una confianza del 95% es:

Límite inferior: $(O/E) \cdot \left[1 - (1/9 \cdot O) - (1{,}96/3 \cdot \sqrt{O})\right]^3$

Límite superior: $\frac{O+1}{E} \cdot \left[1 - \left(1/9 \cdot (O+1)\right) + 1{,}96/3 \cdot \sqrt{O+1}\right]^3$

Si el intervalo de confianza de la ratio estandarizada contiene el valor 1, no hay diferencia significativa entre el centro en estudio y el estándar de referencia

Aunque eventualmente calcularemos los dos límites, inferior y superior, del intervalo de confianza, lo que nos interesará en primer lugar es ver si las ratios mayores de 1 son significativas, calculando el límite inferior, o si las menores de

1 lo son, para lo cual calcularemos el límite superior. Veamos los ejemplos de H_1 y H_2, de la Tabla 18.6:

Para H_1: , y calcularemos el límite superior para ver si está por debajo de 1, lo cual significaría que la tasa en el H_1 es significativamente menor que la esperada:

$$\text{Límite superior} = \left(\frac{45}{68,9}\right) \cdot \left[1 - \left(1/9 \cdot (45)\right) + \left(1,96/3 \cdot \sqrt{45}\right)\right]^3 = 0,75$$

Al ser menor de 1, concluiremos que la tasa de cesáreas en el hospital 1 es significativamente menor que el estándar nacional.

Para H_2, con una ratio = 1,08, nos interesa el límite inferior, que tendrá que ser mayor de 1 para concluir que la tasa en el hospital 2 es significativamente mayor que la estándar. Los cálculos serían:

$$\text{Límite inferior} = (1,08) \cdot \left[\left(1 - (1/9 \cdot (150)\right) - \left((1,96/3 \cdot \sqrt{150})\right)\right]^3 = 0.91$$

No podemos concluir que la tasa en H_2 sea significativamente mayor que la estándar porque el límite inferior del intervalo de confianza es menor que 1 y por lo tanto el valor 1 está contenido en este intervalo.

Los procedimientos para ver la significación estadística de las comparaciones con el estándar basadas en el cálculo del valor de z, o del intervalo de confianza simple de la ratio estandarizada, no tienen en cuenta la posible variabilidad dentro de los estratos que resume el indicador

Tanto el procedimiento que acabamos de ver, como el cálculo del valor de z para la diferencia entre casos observados y esperados, tienen en cuenta exclusivamente la variabilidad y consiguiente probabilidad de error, debida al muestreo en general. La modificación introducida por Lawthers, Palmer, Edwards et al, consiste en la aplicación de un modelo jerárquico simple que tiene en cuenta también la variabilidad dentro de cada uno de los estratos relevantes del factor por el que se haya ajustado el indicador. Por ejemplo, si el indicador se ha ajustado por niveles de riesgo, se tiene en cuenta la variabilidad dentro de cada nivel de riesgo; si se ha ajustado por profesional, o, por ejemplo, por centro, al calcular el indicador de un área geográfica, se tiene en cuenta la variabilidad dentro de cada profesional o dentro de cada centro. El modelo es muy sencillo de aplicar, tanto más cuantos menos estratos y factores se hayan considerado, y al parecer es estadísticamente más aconsejable que los otros procedimientos que acabamos de ver.

d) Comparación de ratios en base a un modelo jerárquico

Lo que haremos es, como en el procedimiento anterior, calcular los límites del intervalo de confianza de la ratio estandarizada, sólo que teniendo en cuenta al menos dos niveles de variabilidad:

1. La correspondiente a la muestra general, representada por los valores observados (O).
2. La correspondiente a los estratos del factor por el que se ha ajustado, representados por los valores Oi, siendo i el número de estratos. Por ejemplo si se ha ajustado por tres niveles de gravedad, tendremos los valores observados O_1, O_2, y O_3, correspondientes a los tres estratos o categorías del factor por el que ajustamos.

El modelo muestra que la varianza total, bajo ciertas asunciones, está representada por $V = O + \Sigma (Oi)^2$ donde O es el valor observado general, y Oi los valores observados en cada estrato. Por ejemplo, para un factor con tres categorías o estratos, la varianza sería: $V = O + (O_1)^2 + (O_2)^2 + (O_3)^2$ y los límites del intervalo de confianza de 95%:

— Límite inferior: $(O/E) - \left[1{,}96 \cdot \sqrt{V/E}\right]$

— Límite superior: $(O/E) + \left[1{,}96 \cdot \sqrt{V/E}\right]$

El intervalo de confianza de la ratio estandarizada puede calcularse en base a un modelo jerárquico que tiene en cuenta dos niveles de variabilidad.

Estos cálculos normalmente amplían el intervalo de confianza, precisando de ratios más alejadas de 1, y/o tamaños de muestra mayores, para que aparezcan valores significativos, pero las conclusiones son más sólidas. Vamos a aplicarlo a un ejemplo. Supongamos que las tasas observadas de cesáreas se han ajustado en base a una escala de riesgo con tres niveles y que los valores observados en H1 son: $O_1 = 12$ $O_2 = 15$ $O_3 = 17$ Total = 44

La varianza total, teniendo en cuenta la varianza dentro de los estratos sería:

$V = 44 + (12)^2 + (15)^2 + (17)^2 = 702$

El límite superior del intervalo de confianza: $0{,}64 + \left[1{,}96 \cdot \sqrt{702/68{,}9}\right] = 1{,}39$

Según este resultado, teniendo en cuenta la variabilidad intraestrato, este hospital no tendría una tasa de cesáreas significativamente más baja que el estándar.

Aparte de la comparación de cada centro con el estándar, puede interesarnos averiguar si hay diferencia significativas entre los centros, e incluso fijar un estándar empírico diferente al promedio nacional, basándonos sólo en los datos de los centros de nuestra región o área geográfica, estándar que puede ser también de tipo valor promedio o de otra naturaleza.

Ambos procedimientos están metodológicamente relacionados y vamos a ver a continuación como pueden realizarse.

6.2. COMPARACIÓN DE LOS RESULTADOS DE LOS CENTROS ENTRE SÍ

Para analizar la variabilidad de los centros entre sí existen diversos métodos, de los cuales detallamos uno de los más sencillos.

Es para este objetivo que se han desarrollado metodologías diversas en relación al análisis de la variabilidad en áreas geográficas pequeñas, teniendo como unidad de análisis un área geográfica determinada (distrito, área, etc.) y viendo diferencias para indicadores poblacionales. Más adelante daremos cuenta resumida de los diversos métodos y fórmulas utilizados, que pueden ser perfectamente adaptados al análisis de los resultados de la monitorización externa de indicadores. Pero veremos en detalle uno de los más sencillos, a la vez que estadísticamente suficientemente sólido, relacionado con el método de comparación entre valores observados y esperados que hemos utilizado para averiguar las diferen-

cias con el estándar. Los sucesivos pasos a realizar están reflejados en la Tabla 18.8. Para ilustrar el procedimiento, vamos a volver al ejemplo de la Tabla 18.6.

TABLA 18.8. Pasos a realizar para la comparación de un indicador entre un grupo de centros

PASO	PROCEDIMIENTO
1. Calcular la ratio estandarizada del grupo.	$\Sigma O_i / \Sigma E_i = R_i$
2. Cálculo estandarizado, por la ratio del grupo, de los valores esperados en cada centro.	$(E_i) \cdot R_i = E_i'$
3. Cálculo para cada centro de las diferencias y ratios entre valores observados y los valores esperados calculados en el paso 2.	$O_i - E_i'$ $\quad O_i / E_i'$
4. Cálculo de las desviaciones estándar y valores Z para las nuevas tasas esperadas.	$S_i' = \sqrt{n_i \cdot p'_i \cdot (1 - p'_i)}$ $\quad Z_i = \frac{O_i - E_i'}{S_i'}$
5. Identificar los centros con valores significativamente diferentes.	Ver valor de Z para α/k
6. (Opcional) Ver grado de homogeneidad del grupo.	Calcular la desviación estándar de los valores de Z, y buscar en las tablas de χ^2 el valor correspondiente a la significación deseada y k-1 grados de libertad. $\chi^2 = S_Z^2 \cdot (k-1) \rightarrow S_Z = \sqrt{\frac{\chi^2}{k-1}}$

O = valores observados; E= Valores esperados; ´: indica cálculos de estandarización interna. k= número de centros que se comparan; Si´: desviación estándar de la tasa esperada; Sz: Desviación estándar de los valores de z.

El primer paso para la comparación intragrupo es calcular el estándar del grupo, representado por la ratio del total de casos observados sobre el total de casos esperados.

- **Paso 1. Cálculo de la ratio estandarizada del grupo**. La primera tarea a realizar para comparar los hospitales entre sí sería establecer un estándar general que los represente, especie de estándar "interno" para el grupo de hospitales, con respecto al cual vamos a compararlos de nuevo. Este estándar del grupo se establece en base a los casos observados y esperados de todos los centros en su conjunto, dividiendo el total de casos observados por el total de casos esperados. En nuestro ejemplo esta ratio es:

$$\frac{\Sigma O_i}{\Sigma E_i} = \frac{44 + 150 + 35 + 38 + 35}{68{,}9 + 138{,}3 + 36{,}7 + 56{,}4 + 50{,}9} = 0{,}86$$

- **Paso 2. Estandarización por la ratio del grupo de los valores esperados en cada centro**. Tras esta operación, multiplicamos los valores esperados de cada centro por la ratio del grupo; los valores esperados resultantes (E´i) deberían ser iguales a los observados, de forma que las variaciones entre centros serán debidas a su propia diferenciación de cada centro con respecto al grupo. El

resultado de esta estandarización intragrupo, así como de los cálculos que se han de realizar a continuación, están reflejados en la Tabla 18.9.

La ratio del grupo se multiplica por los valores esperados de cada centro para obtener nuevos valores esperados en función del estándar del grupo.

TABLA 18.9. Tasa de cesáreas en cinco hospitales. Resultados del ajuste y comparación intragrupo

n	O	E'	O-E'	O/E'	DE	z
288	44	59,3	-15,3	0,74	6,86	-2,23
585	150	118,9	31,1	1,26	9,73	3,20
141	35	31,6	3,4	1,11	4,95	0,69
269	38	48,5	-10,5	0,78	6,31	-1,67
225	35	43,8	-8,8	0,80	5,94	-1,48
	302	302			Media	-0,30
					Desv. Estándar	2,25

E´: Valores esperados ajustados por la ratio estandarizada del grupo.

- **Paso 3. Cálculo de las diferencias y ratios entre valores observados y los valores esperados estandarizados por la ratio del grupo**. Con estos valores de O y E´ podemos proceder de la misma forma que vimos en el apartado anterior para comparar cada centro al estándar promedio nacional viendo la diferencia y ratio para cada centro.
- **Paso 4. Cálculo de las desviaciones estándar y valor z para las nuevas tasas esperadas**. La forma de calcularlos es idéntica a la que vimos al comparar con el estándar promedio nacional. Por ejemplo para H_1:

$$DE = \sqrt{288 \,.\, (59{,}3/288)\,(1 - 59{,}3/288)} = 6{,}86$$

$$Z = \frac{-15{,}3}{6{,}86} = -2{,}23$$

De igual manera lo calcularemos para los otros cuatro centros.

- **Paso 5. Identificar los centros con valores significativamente diferentes**. Para evitar señalar significaciones estadísticas ≤ 0,05 que aparezcan por azar, y teniendo en cuenta que realizamos cinco comparaciones establecemos como significativa una p<0,01 y una z/2,58 (ver Tabla 18.7). Según esta regla, resulta significativamente mayor la tasa del hospital 2, con una z=3,2. Lo lógico sería entonces investigar cuales son las causas de que en ese centro tengan un 26% más de cesáreas que en el estándar regional, incluso ajustado por riesgo.

Para ver qué centros tienen diferencias significativas en relación al ratio estándar del grupo se calculan las desviaciones estándar y el valor de z para las tasas esperadas que hemos obtenido al estandarizar por la ratio del grupo.

- **Paso 6. Comprobar el grado de homogeneidad del grupo**. Aunque los objetivos de nuestro análisis puede que los demos por concluidos una vez identificados los centros sobre los que hay que intervenir, podemos adicionalmente comprobar si los resultados entre los centros son globalmente homogéneos. En realidad, esta comprobación puede realizarse como paso previo a identi-

El grado de homogeneidad del grupo puede averiguarse en función del valor de la desviación estándar de los valores de z.

El producto de la varianza de los valores de z y el número de centros que se comparan menos uno (k−1), sigue una distribución de probabilidad χ^2 con k−1 grados de libertad.

ficar los centros con valores significativamente diferentes, por cuanto si nos resultan homogéneos no es preciso investigar diferencias. Por otra parte, es a través de comprobaciones sucesivas de la homogeneidad como podemos llegar a identificar un estándar de mejor práctica (excelencia relativa) realista y propio del grupo que consideremos, tal como veremos más adelante. La homogeneidad puede comprobarse buscando el valor límite permisible, al nivel de significación deseado, para la desviación estándar de los valores z del grupo de centros. Este valor límite se busca en las tablas de la distribución chi cuadrado (χ^2), que es la que sigue la fórmula $S_z^2 . (k-1) = \chi^2$, donde S_z es la desviación estándar de los valores z, y k es el número de centros que se comparan. Según esta fórmula calcularemos el valor máximo de la desviación estándar de z para el nivel de significación que elijamos.

En nuestro ejemplo de la Tabla 18.9, buscamos en las tablas el valor χ^2 de para p<0,01 (la significación que queremos establecer) y 4 grados de libertad (k-1), que resulta ser = 13,2767 a cuyo valor le corresponde un valor máximo de z, que calculamos en base a la fórmula vista más arriba:

$$S_z^2 . (k-1) = \chi^2 \rightarrow S_z = \sqrt{\frac{\chi^2}{k-1}} = \sqrt{\frac{13{,}2767}{4}} = 1{,}82$$

Según este resultado, si la desviación estándar de Z es ≥ 1,82 concluiremos que los centros no tienen un comportamiento homogéneo, para p<0,01. En nuestro ejemplo (Tabla 18.9) Sz=2,25, con lo cual concluiremos que los centros no son homogéneos; ya sabemos, además, que H_2 tienen una tasa significativamente mayor.

7. IDENTIFICACIÓN DE ESTÁNDARES DE EXCELENCIA RELATIVA EMPÍRICOS Y REALISTAS

Puede resultar de interés para la gestión de la calidad identificar el estándar de excelencia relativa, propio del grupo.

Hasta aquí hemos visto el procedimiento más común en la monitorización externa de indicadores, que es comparar cada centro o unidad de análisis con el estándar que representa el promedio nacional o regional. También hemos visto cómo comparar grupos de centros entre sí, teniendo en cuenta su propio estándar promedio. Sin embargo, para fijar objetivos a corto y medio plazo, y también para realizar comparaciones y análisis más en línea con la gestión de la calidad, puede interesarnos identificar en el grupo de centros considerado, un estándar de excelencia relativa, que podemos definir como el nivel más elevado que se ha alcanzando de forma consistente dentro de nuestro grupo de centros. Como vamos a ver, no se trata de seleccionar el centro con los mejores resultados, que podría ser visto como caso excepcional por circunstancias diversas, sino de encontrar el mejor resultado que puede obtenerse de forma homogénea, una vez eliminados los centros que lo hacen significativamente peor que la media, e incluso los casos excepcionalmente buenos. El procedimiento se basa en tests de homogeneidad sucesivos, hasta que no aparezca ningún centro con una tasa significativamente diferente como para que el conjunto de centros no resulte homogéneo. Este grupo final de centros con tasas homogéneos representa la excelencia relativa del grupo. Estamos hablando de centros, pero el procedimiento puede aplicarse igualmente a grupos de otro tipo de unidades de servicio, incluso profesionales individuales dentro de un mismo centro, siempre que estemos tratando de un mismo tipo de servicios o pacientes.

La excelencia relativa se refiere al mejor resultado que puede obtenerse de forma homogénea, una vez eliminados los centros significativamente diferentes.

Para ilustrar la forma de proceder, vamos a volver a nuestro ejemplo de la tasa de cesáreas, Tablas 18.8 y 18.9.

En base a los cálculos realizados para la comparación intragrupo (Tabla 18.9) hemos identificado que los resultados no son homogéneos (el valor de la desviación estándar de los valores de z es mayor que el límite permitido para p<0,01), y que uno de los centros (H_2) tiene una tasa significativamente mayor (nota: los datos de la Tabla 18.8 podrían también ser utilizados si el estándar utilizado para comparar hubiese sido el promedio de los cinco centros, en vez del promedio nacional; y el mismo procedimiento que vamos a ilustrar podría valorarse a nivel regional o nacional utilizando inicialmente los correspondientes estándares promedios como referencia). Para buscar la tasa homogénea, eliminamos los datos de H_2 y recalculamos, los datos para los cuatro centros restantes, ajustando esta vez por la ratio O/E conjunta de estos cuatro centros solamente. Esta nueva tasa conjunta $\Sigma O_i/\Sigma E_i$ es 0,714. El resultado de los nuevos cálculos para los cuatro centros que estamos considerando se refleja en la Tabla 18.10.

El procedimiento para identificar el estándar de excelencia relativa del grupo se basa en tests de homogeneidad sucesivos, eliminando en cada paso los centros significativamente diferentes, hasta conseguir la homogeneidad.

TABLA 18.10. Resultados de los cálculos para la identificación de un estandar de excelencia relativa

n	O	E'	O-E'	O/E'	DE	z
288	44	49,19	-5,2	0,89	6,39	-0,81
141	35	26,20	8,8	1,34	4,62	1,90
269	38	40,27	-2,3	0,94	5,85	-0,39
225	35	36,34	-1,3	0,96	5,52	-0,24
	152	152				
					Media	0,09
					Desv. Estándar	1,22

E´: Valores esperados ajustados por la ratio estandarizada del grupo

El valor máximo de Sz para el test de homogeneidad, dado que, según las tablas, χ^2 para 3 grados de libertad y p<0,01 es 11,3449, resulta

$$S_Z = \sqrt{\frac{11,3449}{4}} = 1,68$$

Como el resultado obtenido (1,22) es menor que 1,68, concluimos que las tasas son homogéneas y la tasa promedio de estos cuatro centros sería la tasa de excelencia relativa del grupo. En nuestro ejemplo, con un grupo pequeño, sólo hemos eliminado un centro, pero en grupos más grandes y/o más heterogéneos el estándar de excelencia relativa puede necesitar de varios cálculos sucesivos, hasta que encontremos la homogeneidad que vamos buscando como tasa de excelencia relativa, empírica y realista; que podemos por otra parte ir modificando con el tiempo a medida que vayan mejorando los indicadores en el grupo de centros considerado.

Cuanto más grande y/o más heterogéneo sea el grupo, serán necesarios más análisis de homogeneidad sucesivos para identificar el estándar de excelencia relativa.

8. ANÁLISIS GRÁFICO DE LA MONITORIZACIÓN EXTERNA

El análisis gráfico de la monitorización externa de indicadores emplea las mismas técnicas que vimos para presentar los datos de una evaluación.

El análisis gráfico de los datos de una monitorización externa no difiere, adaptándolo, del que vimos en la UT 11 para presentar los datos de una evaluación. En el caso de la monitorización de un indicador en varios centros, cada uno de los centros puede ser considerado a efectos de representación grafica como cada uno de los criterios de calidad que veíamos en la UT 11, y representar los resultados de la misma forma que se representarían los resultados de la evaluación de varios criterios.

Así, la forma más frecuente de presentar los datos comparativos de varios centros es con un gráfico de barras, mejor si se ordenan de mayor a menor, en el que cada barra representa un centro y al que se suele añadir la barra correspondiente al promedio o estándar de referencia.

Es también muy útil y frecuente la utilización de gráficos tipo box and whiskers, tanto para representar los valores de centros individuales (como en la Figura 18.1), como grupos de centros.

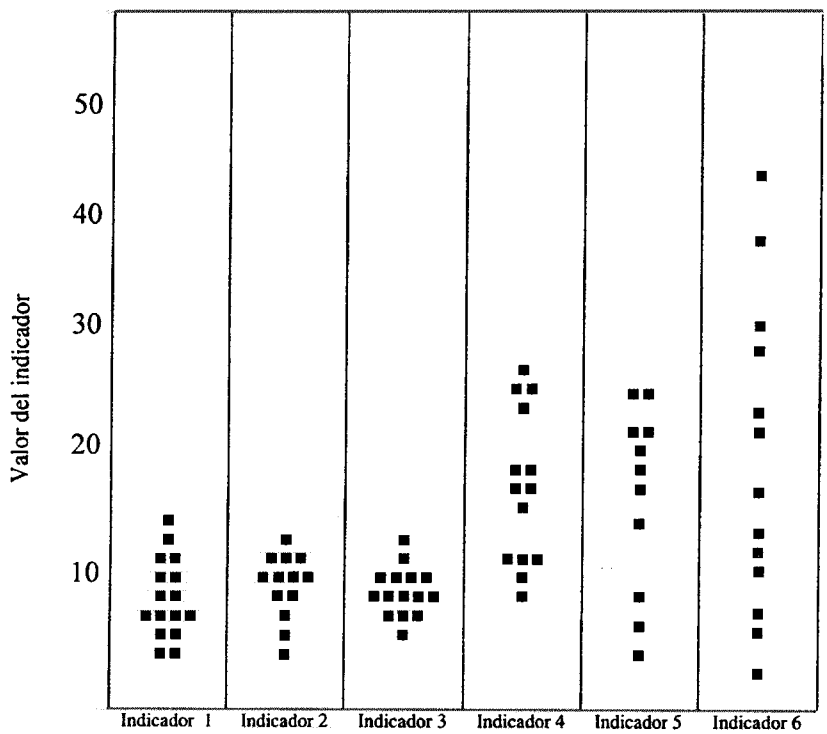
Son especialmente útiles los gráficos de barras (ordenadas), los de box and whiskers, y los de puntos agrupados según diversos indicadores.

Finalmente, cuando lo que interesa es representar en el mismo gráfico tres variables como pueden ser: (i) el nivel de cumplimiento del indicador; (ii) cada uno de los centros; y (iii) varios indicadores a la vez, se emplean gráficos de puntos agrupados como los de la Figura 18.2, en los cuales el nivel de cumplimiento de los indicadores se representa en el eje de ordenadas, los diversos indicadores (o categorías de la variable que nos interese) en el de abscisas, y cada centro como un punto en cada uno de los grupos de puntos que corresponde a cada uno de los indicadores o categorías que figuran en el eje de abscisas.

Cualquiera de los métodos de análisis gráficos que se utilicen, deben tener como objetivos:

1. Visualizar la variabilidad.
2. Visualizar la comparación de cada centro con el estándar.

FIGURA 18.2. Gráfico de puntos agrupados para comparar la variabilidad de varios indicadores.



■ : *cada punto es un centro o unidad de servicio de interés*

9. OTROS MÉTODOS DE ANÁLISIS DE LA VARIABILIDAD DE LOS RESULTADOS DE UN INDICADOR

El tipo de análisis de los datos comparativos de un indicador en varios centros o unidades que hemos visto en esta UT de análisis es fácilmente ejecutable y sirve a los objetivos de la gestión de la calidad (identificar centros con un problema potencial de calidad, en comparación al estándar que hayamos definido). Pero también hemos mencionado que las metodologías desarrolladas para analizar la variabilidad de determinadas tasas de utilización de recursos entre pequeñas áreas geográficas (*Small Area Analysis*, iniciado por Wennberg y Gitterson a principios de los 70), serían perfectamente adaptables al análisis también de la variabilidad por centros de los resultados de determinados indicadores. Estos métodos de análisis, sin embargo, se han centrado más en la detección de la variabilidad estadísticamente significativa en el grupo de áreas analizadas y su posible relación con determinadas causas hipotéticas, que en determinar cuánta variación es admisible, o cuál sería el estándar adecuado para comparar. Adicionalmente, son sorprendentemente frecuentes los estudios publicados que utilizan el método indirecto para estandarizar las tasas; método que, como hemos visto, sólo serviría para comparar cada área con el estándar (normalmente la media regional o nacional) pero no con otras áreas. No obstante, se han ido desarrollando diversas fórmulas, cada vez más refinadas y robustas (en el sentido de tener en cuenta diversas circunstancias y ser aplicables con cada vez menos condicionantes y restricciones) que resumimos en la Tabla 18.11, todas ellas con una referencia a la que se puede acudir en caso de estar interesados en los detalles de su cálculo y aplicación.

TABLA 18.11. Métodos para comparar variabilidad en pequeñas áreas geográficas (Small Area Analysis)

MÉTODO	OBSERVACIONES
1. Ratio de tasas extremas. (tasa más alta/tasa más baja)	• Es el más simple, pero también el más desacreditado.
2. Coeficiente de variación. (desviación estándar de las tasas/tasa promedio)	• No tiene en cuenta la variabilidad intra-área.
3. Coeficiente de variación ponderado.	• Coeficiente de variación ajustado por las diferencias en el tamaño de la población de las áreas.
4. Coeficiente de variación "verdadero".	• Coeficiente de variación basado en el análisis de varianza entre y dentro de las áreas.
5. Componente sistemático de la varianza.	• Estima el coeficiente de variación, eliminando la variación intra-áreas.
6. χ^2 de valores esperados vs observados	• Aplicable, con las debidas precauciones en cuanto a significación estadística, tanto para valoración global como para cada área.
7. T2, una fórmula cuyos resultados siguen la distribución χ^2 de probabilidad.	• Test robusto, aplicable para comprobar variaciones en tasas de baja incidencia.
8. Modelo logístico jerarquizado.	• Modela el análisis de la variabilidad en tres niveles (variación intra-área, variación entre áreas, modelo general) • No precisa estandarización de tasas; los factores de ajuste se incluyen en el modelo de regresión logística.

REFERENCIAS:
- Una explicación comparativa de los métodos 2, 3, 4 y 5 puede verse en: Diehr P, Cain K, Ye Z, Abdul-Salam F. Small Area Variation Analysis. Methods for Comparing Several Diagnosis-Related Groups. Med Care 1993; 31(5): Y545-Y553. Supplement
- Una aplicación del método 6 puede verse en: Connell FA, Day RW, Logerfo JP. Hospitalization of Medicaid Children: Analysis of Small Area Variations in Admission Rates. Am J Public Health 1981; 71(6):606-13.
- El método 7 se explica en: Carriere KC, Roos LL. A method of Comparison for Standarized Rates of Low-Incidence Events. Med Care 1997; 35(1): 57-59.
- El método 8 se explica en: Gatsonis C, Normand S-L, Lin C, Morris C. Geografic Variation of Procedure Utilization. A Hierarchical model Approach. Med Care 1993; 31(5): Y554-Y559, Supplement.

BIBLIOGRAFÍA

- Goldfield N, Pine M, Pine J. Measuring and Managing Health Care Quality. Aspen: Gaithersburg, MD; 1995.
- Orav EJ. Statistitcal Issues for Rate-based Measurement. En: Schoenbaum SC, Sundwell ON, Bergman et al. Using Clinical Practice Guidelines to Evaluate Quality of Care. U.S. Washington: Department of Health and Human Services. AHCPR Pub. 95-0046; 1995.
- Lawthers A, Palmer RH, Edwards JE et al. Developing and evaluating performance measures for ambulatory care quality. The Joint Commission on Quality Improvement; 1993. 16(12): 552-65.
- Rothman KJ. Modern Epidemiology. Boston: Little, Brown and Co.; 1986.

UNIDAD TEMÁTICA 19

19

INTRODUCCIÓN A LA PLANIFICACIÓN O DISEÑO DE LA CALIDAD

EMCA

Gestión de la Calidad Asistencial

CONTENIDO GENERAL

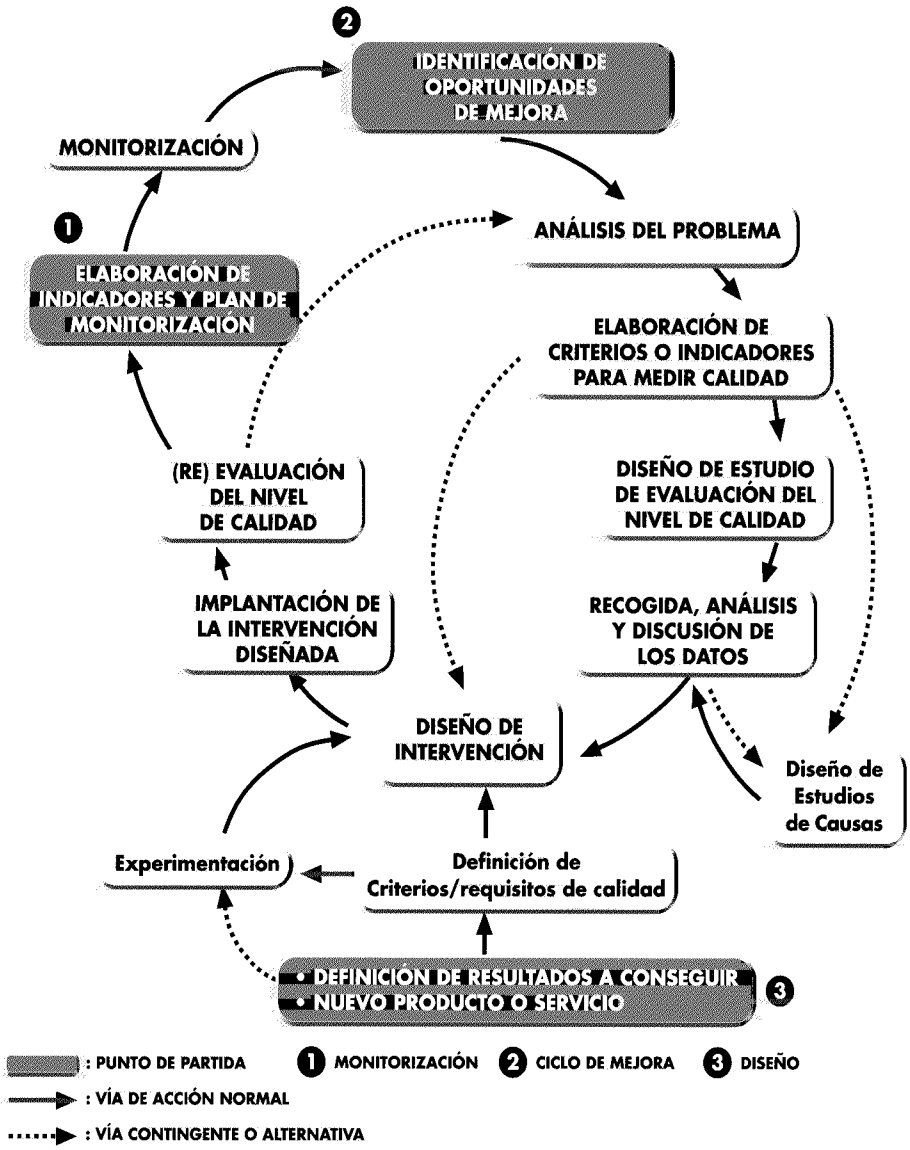
Esta UT ofrece una introducción a los principales conceptos y esquemas metodológicos del diseño de la calidad. Con ello se completa la visión panorámica de los métodos y actividades a nuestro alcance para practicar nuestro compromiso con la calidad en los servicios de salud, a la vez que se anuncian las futuras vías de formación que pueden recorrer aquellos que se interesen por liderar la práctica del compromiso con la calidad en sus respectivos entornos de trabajo.

ÍNDICE DE CONTENIDOS

1. Introducción.
2. Diseño de la calidad: Concepto e importancia.
3. Puntos de partida para el diseño de la calidad.
4. Esquema metodológico básico.
5. La protocolización como una actividad de diseño de la calidad.
6. Evaluación de la calidad de los protocolos clínicos.
7. Diseño total de los servicios: la última generación de protocolos.
8. Métodos de mejora de la calidad con base en el diseño.

OBJETIVOS ESPECÍFICOS

1. Describir el esquema metodológico básico de las actividades de diseño o planificación de la calidad.
2. Enumerar diversas metodologías utilizables en la planificación o diseño de la calidad.
3. Explicar y analizar la protocolización de procesos asistenciales como actividades de diseño de la calidad.
4. Determinar las necesidades de formación futuras en estos temas.



1. INTRODUCCIÓN

En el caso del diseño o planificación de la calidad, el grupo de actividades al que más importancia se le está dando últimamente, vamos a ver cuál es el esquema metodológico básico en el que se enmarcan cualquiera de los muchos métodos y enfoques existentes, y haremos un cierto énfasis en el concepto e implicaciones de la protocolización de actividades como método de diseño de la calidad.

Terminamos la UT reflexionando sobre lo que queda por hacer. El camino de la calidad, como el de la ciencia, no tiene fin ni verdad absoluta, pero tampoco vuelta atrás.

En esta UT se da una visión panorámica de los métodos y actividades de monitorización diseño de la calidad

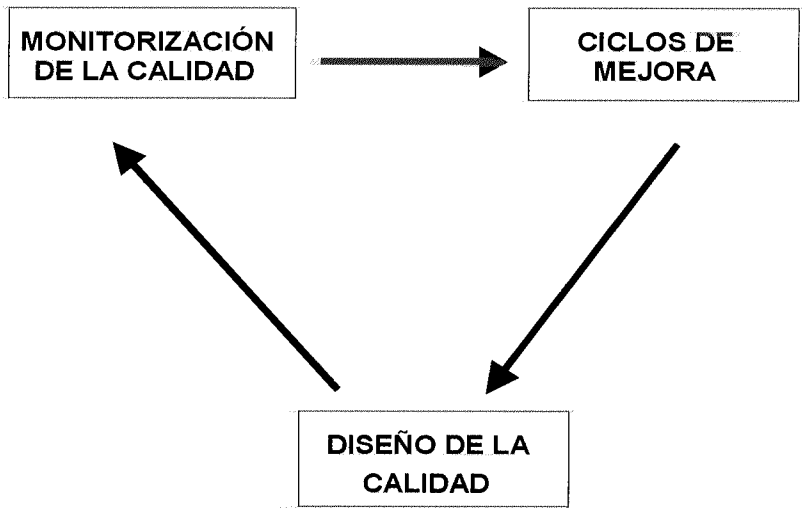
2. DISEÑO DE LA CALIDAD: CONCEPTO E IMPORTANCIA

El diseño de la calidad sería definido por muchos como el tercer y más importante componente de los programas de gestión de calidad. Es curioso como se han ido originando diversos esquemas que distinguen justamente tres tipos de actividades y metodologías dentro de los programas de gestión de calidad. Así, por ejemplo, la "trilogía" de Juran incluye *planificación*, control y mejora de la calidad; el "sistema de control total de la calidad" de Feigenbaum implica *desarrollo de calidad*, mantenimiento de la calidad y mejora de la calidad; Ishikawa habla de los "tres métodos de garantía de calidad": inspección, control de procesos y *énfasis de las innovaciones y desarrollo de nuevos productos*; H. Palmer describe evaluaciones retrospectivas, concurrentes y *prospectivas* de la calidad. En todos estos casos, como en nuestro esquema de ciclos de mejora, monitorización y diseño de la calidad, el componente subrayado se refiere a predecir los procesos y conseguir que se tomen las decisiones correctas para lograr unos resultados predeterminados; en definitiva, prevenir la aparición de problemas de calidad (Tabla 19.1 y Figura 19.1).

TABLA 19.1. Grupos de actividades de los programas de Gestión de la Calidad. Puntos de partida normales y objetivos inmediatos

GRUPO DE ACTIVIDADES	PUNTO DE PARTIDA	OBJETIVO INMEDIATO
Ciclos de mejora	• Identificación de un problema de calidad u oportunidad de mejora en algún aspecto de los servicios que se ofrecen	• Solucionar el problema • Aprovechar la oportunidad de mejora descubierta
Monitorización	• 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.	• Identificación de problemas de calidad u oportunidades de mejora
Diseño	• Programación de un nuevo servicio a ofrecer. • Identificación de necesidades y expectativas de los usuarios. • Identificación de parámetros y resultados a conseguir	• Diseñar los procesos de atención para conseguir los resultados deseados predeterminados

FIGURA 19.1. Los grupos de actividades de los Programas de Gestión de la Calidad.



Aplicado a los servicios de salud, una definición más detallada de lo que entendemos por diseño de la calidad sería la siguiente: *Saber lo que hay que hacer para solucionar los problemas de salud y satisfacer las expectativas de los usuarios, y poner los medios y organizar los procesos de manera que sea lógico, fácil e inevitable ofrecer un servicio con la calidad esperada.*

Diseñar la calidad es prevenir la aparición de problemas y garantizar unos resultados predeterminados

3. PUNTOS DE PARTIDA PARA EL DISEÑO DE LA CALIDAD

Las actividades de diseño de la calidad pueden iniciarse a partir de tres situaciones diferentes:

1. Al diseñar intervenciones en los ciclos de mejora.
2. Para introducir mejoras de forma planificada en los productos o servicios que ofrecemos, innovaciones que se establecen no sobre la base de la manera en que se ofrecen los servicios habitualmente (como en los ciclos de mejora) sino introduciendo objetivos o aspectos nuevos.
3. Al diseñar nuevos servicios, previamente inexistentes.

En todos los casos, la base común es que partimos de la definición de unos resultados a conseguir, y de ahí intentamos averiguar cómo diseñar los procesos para conseguirlos. Asegurarse los resultados deseados cuando se consuma el producto o servicio diseñado, es de gran relevancia sobre todo en el caso de los servicios; tanta que es explicable el énfasis cada vez mayor en este componente de los programas de gestión de la calidad. Las principales razones se resumen en la Tabla 19.2.

El diseño de la calidad es de especial importancia en los servicios y entre ellos en los servicios de salud.

TABLA 19.2. Importancia del diseño en la calidad de los servicios de salud

- EL SERVICIO SE FABRICA Y SE CONSUME AL MISMO TIEMPO
- EL CLIENTE / USUARIO NO PUEDE PROBAR EL SERVICIO ANTES DE ADQUIRIRLO.
- UN SERVICIO DEFECTUOSO NO PUEDE REPARARSE, NI REVENDERSE A BAJO PRECIO. ES UNA PERDIDA IRREPARABLE.
- LOS SERVICIOS IMPLICAN MÚLTIPLES CONTACTOS QUE AUMENTAN LAS POSIBILIDADES DE EQUIVOCARSE.

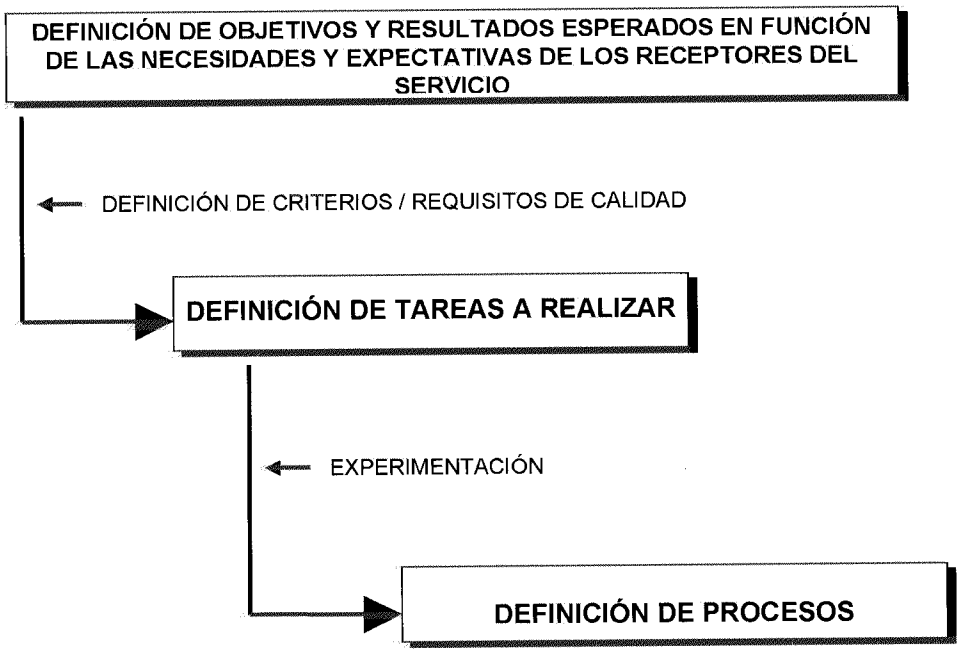
ES, POR TANTO, ESPECIALMENTE IMPORTANTE DISEÑAR LOS SERVICIOS DE SALUD PARA HACERLO BIEN SIEMPRE A LA PRIMERA

Adaptado de Horovitz J. La calidad del servicio.

4. ESQUEMA METODOLÓGICO BÁSICO

Las actividades y métodos que implican o se relacionan con el diseño de la calidad pueden ser muy complejos, de hecho los más complejos de todo el programa de gestión de la calidad. Sin embargo en sus rasgos comunes, esenciales, pueden resumirse en el esquema de la Figura 19.2.

FIGURA 19.2. Diseño de la calidad. Esquema básico



Este esquema implica lo siguiente:

1. El diseño comienza por tener claros los *objetivos y resultados esperados*, en función de las necesidades y expectativas de aquellos para quienes va a estar pensado el servicio (sus receptores, usuarios o clientes potenciales). Para ello

hay que manejar las metodologías conducentes a conocer necesidades y expectativas de los potenciales usuarios, e identificar qué quiere decir que los resultados sean satisfactorios, lo cual nos ha de conducir a *identificar las características, requisitos o criterios* asociados al resultado esperado.

2. El siguiente componente sería identificar qué *tareas* o *actividades* se han de realizar para conseguir esos criterios o requisitos de calidad.

   Frente a las diversas opciones, lo normal sería *experimentar* o pilotar diversas alternativas o combinaciones, de forma que tengamos base empírica y científica para el tercer y definitivo componente del diseño.

3. Definir (diseñar) los *procesos y métodos concretos de funcionamiento* que nos van a garantizar los resultados deseados, que fueron definidos como punto de partida, cuando se tenga la ocasión de ofrecer el servicio diseñado.

Las actividades y métodos para el diseño de la calidad son numerosos, aunque su utilización responde a un esquema básico en el que se parte del conocimiento de necesidades y expectativas de los receptores de los servicios y se termina con el diseño de los procesos para conseguir hacerlo bien a la primera y siempre que se ofrezca el servicio diseñado.

Si tuviésemos que mencionar las principales áreas sustantivas en cuanto a métodos que intervienen o pueden intervenir en alguna de las fases de diseño, destacaríamos aquellas en relación con la investigación de necesidades y expectativas del usuario, la investigación de servicios de salud, las técnicas de investigación operativa, el diseño de experimentos y las técnicas de valoración y resumen de la evaluación científica. Muchos de estos métodos y técnicas, no todos, se han desarrollado completamente en la industria, de la que podemos aprender y adaptar un número considerable de ideas y procedimientos para el diseño de servicios de salud. Sin embargo, la forma más tradicional de diseño de la calidad en los servicios de salud es probablemente la protocolización de actividades, incluidos, y sobre todo, los *protocolos clínicos* llamados últimamente *guías de práctica clínica*.

5. LA PROTOCOLIZACIÓN COMO UNA ACTIVIDAD DE DISEÑO DE LA CALIDAD

La relevancia y el desarrollo creciente que están teniendo los programas de gestión de calidad en todos los sectores, incluido el sanitario, y el impulso que desde esta óptica se le está dando a la protocolización, hace conveniente explicitar y especificar el lugar de la protocolización en el conjunto de actividades para la mejora de la calidad.

La protocolización es, con todas sus consecuencias metodológicas, una actividad de diseño, planificación o desarrollo de la calidad. Para diseñar calidad se ha de partir, como hemos apuntado, de la definición de los resultados que se quieren conseguir en relación a las necesidades y expectativas del cliente para el que está pensado el producto o servicio. Sobre esta base, es decir, una vez definido el cliente, necesidad a satisfacer y resultados que se quieren conseguir, se analiza y utiliza toda la evidencia existente y/o se investiga y subrayan los puntos no suficientemente conocidos, para terminar diseñando las características concretas del producto o del servicio en consideración de manera que, cuando se consuma, produzca en sus usuarios el resultado deseado. Como todo diseño, anticipándose a los posibles problemas, el protocolo debe tratar de conseguir que se hagan las cosas bien a la primera y en todas las ocasiones en que el protocolo sea aplicable; se diseña el proceso de atención de forma que se consiga el mejor resultado posible, aumentando así la calidad de la atención prestada

La protocolización es una actividad de diseño de la calidad. Si lo que se diseña es la actuación clínica estamos ante un protocolo clínico, también conocido últimamente como Guía para la Práctica Clínica.

o, lo que es lo mismo, aumentando la efectividad, la eficiencia y la satisfacción previendo y previniendo los posibles problemas de calidad.

Los protocolos clínicos, bajo cualquiera de los muchos términos con que se le: quiera nombrar (el último y más extendido: guías de práctica clínica, a propues ta del Instituto de Medicina de los Estados Unidos, IOM) se presentan como ins trumentos que facilitan la toma de decisiones, disminuyen la incertidumbre y lc variabilidad de la práctica clínica, mejorando así la calidad de la asistencic prestada. Por esta razón están suscitando un interés creciente, y no sólo poi parte de los proveedores de atención sanitaria para su práctica clínica habitual sino también entre los legisladores, planificadores y gestores de servicios de sa· lud. Interés que se justifica no sólo por la repercusión de los protocolos en lc calidad de la atención prestada y el control de los riesgos de ella derivados, sinc también por su efecto en el control de los costes y el uso inadecuado de lo: recursos, y en la reducción del riesgo de posibles reclamaciones legales.

A pesar de todo ello, la protocolización y los protocolos clínicos no son aún ur tema exento de polémica, relacionada generalmente con una serie de factore: entre los que destaca la ausencia de un claro, adecuado y aceptado marco con· ceptual y metodológico. Algunos síntomas de esta situación son, por ejemplo, lc existencia de una cierta confusión terminológica y el considerar, por extensión, los defectos que puedan tener algunos malos protocolos (falta de flexibilidad, validez y aplicabilidad cuestionable, etc.), como defectos intrínsecos a la proto· colización en sí.

5.1. ¿QUÉ ES UN PROTOCOLO CLÍNICO?

Una de las circunstancias que más influye en la confusión terminológica sobre los protocolos y la protocolización, es la multitud de significados existentes aso· ciados al término de protocolo clínico, así como multitud de términos que respon· den a un mismo concepto. Las diferentes definiciones han sido elaboradas por instituciones, autores aislados o grupos de ellos, y publicadas en medios de difu· sión científicos de diversa índole.

La proliferación de términos, también ha existido y aún existe en otros países como USA, Reino Unido, o Canadá en los que se ha propuesto en un determina· do momento, un término único para englobar los existentes, cuya entidad, sin embargo, se reconoce como términos sinónimos aunque a veces parciales, o referidos a un aspecto o formato de presentación particular. El nuevo término es el propuesto por el IOM: guías para la práctica clínica (GPC, *Clinical Practice Guidelines*). Los términos que se recogen *explícitamente* como sinónimos de GPC, es decir respondiendo a conceptos similares en la propuesta original del IOM y en otras varias como en la iniciativa canadiense, son los de "protocolos" (*proto· cols*), "parámetros de práctica" (*practice parameters*), "algoritmos" (*algorithms*), y "herramientas descriptivas" o conjunto de criterios encadenados (*descriptive tool*). Estos y otros términos, nacidos en contextos, autores o grupos diferentes, van siendo sustituidos paulatinamente por el de guías para la práctica, si bien aún coexisten. La razón de seguir llamándoles "protocolos" en este texto es por· que es éste el término más conocido y con más tradición entre los profesionales de la salud en nuestro país.

Los "protocolos clínicos", "parametros de práctica", "algoritmos" y "herramientas descriptivas" son algunos de los formatos a los que se refiere el IOM con el término de "Guía de Práctica Clínica".

Centrándonos en el concepto, en los últimos años la aportación más clarificadora y ajustada a la práctica ha sido la definición realizada por el IOM, que ha tenido un gran impacto en el ámbito sanitario de prácticamente todo el mundo. Esta definición se realiza a través de los elementos y atributos que debe reunir una GPC o protocolo: "Una exposición de principios o recomendaciones para facilitar la toma de decisiones apropiadas en la atención a los pacientes en situaciones clínicas especificas". En esta línea, recogiendo los componentes más esenciales, pragmáticos y contrastados de las definiciones que existen y sobre la base de las necesidades conceptuales surgidas en la práctica de nuestro grupo de evaluación y mejora de los protocolos clínicos en la Región de Murcia, proponemos la siguiente definición más amplia y explícita: *"Un protocolo o guía para la práctica clínica es un instrumento de diseño de la calidad de la atención que explicita las normas de actuación que ayudan a profesionales y usuarios a decidir de la forma más efectiva, eficiente y satisfactoria posible, frente a problemas específicos de promoción, prevención y restauración de la salud, sirviendo además como guía para la evaluación de la calidad en los casos en los que el protocolo sea aplicable".*

La definición establece tres líneas o características generales: 1ª) Instrumento de diseño de la calidad. 2ª) Facilita la toma de decisiones a profesionales y usuarios frente a problemas específicos. 3ª) Guía para la evaluación de la calidad. De ellas, la segunda es la más ampliamente reconocida. De una manera más detallada, esta definición incluye seis elementos o componentes básicos enunciados en la Tabla 19.3.

Un protocolo clínico es un instrumento de diseño de la calidad que facilita la toma de decisiones y sirve como guía para evaluar la calidad asistencial

TABLA 19.3. Elementos de la definición de protocolo o guía para la práctica clínica.

1. INSTRUMENTO DE DISEÑO DE LA CALIDAD
2. EXPLICITA NORMAS DE ACTUACIÓN
3. AYUDA A LA TOMA DE DECISIONES
4. SOBRE PROBLEMAS DE SALUD ESPECÍFICOS
5. PARA PROFESIONALES Y USUARIOS
6. GUÍA PARA LA EVALUACIÓN

6. EVALUACIÓN DE LA CALIDAD DE LOS PROTOCOLOS CLÍNICOS

Evaluar la calidad de un protocolo como herramienta de diseño de la calidad implica comprobar si están presentes una serie de atributos o requisitos que se explicitan en la Tabla 19.4. Aún siendo todos importantes, vamos a destacar de entre ellos la *validez*, probablemente el principal atributo exigible. Su presencia indica que cuando el protocolo se aplica hay una alta probabilidad de alcanzar los resultados previstos. La validez se comprueba evaluando la evidencia científica que justifica las recomendaciones. Debe estar especificado el método empleado para identificar y revisar las evidencias científicas *cuantitativas* en las que se fundamenta; deben constar las fuentes de información utilizadas; debe existir relación entre la evidencia y las recomendaciones; deben mencionarse *cualitati-*

La calidad de un protocolo o guía para la práctica clínica puede evaluarse comprobando si tiene o no una serie de requisitos, de los cuales el más importante es la validez.

va y *cuantitativamente* los beneficios esperables en salud y los riesgos potencia les para la salud, y debe recoger los costes esperables al aplicar el protocolo. La recomendaciones, por tanto, han de tener en cuenta beneficios, riesgos y coste a la luz de la evidencia científica. La validez se relaciona también con otros do importantes atributos: *flexibilidad* y *aplicabilidad clínica*. Los protocolos que nc se basen en evidencia científica y no sean flexibles en su aplicación no pueder ser considerados buenos protocolos clínicos, y por lo tanto, no deben ser toma dos como ejemplo de lo que es un protocolo.

TABLA 19.4. Características de un buen protocolo clínico y su significado para la evaluación

CARACTERÍSTICA	SIGNIFICADO
Validez	Cuando el Protocolo se siga, deben conducir a los resultados previstos. Puede evaluarse indirectamente considerando la relación entre la evidencia científica y las recomendaciones del Protocolo, y la calidad y la forma de evaluar la evidencia científica que se cite en el mismo.
Fiabilidad/ Reproducibilidad	Con la misma evidencia científica y métodos de desarrollo del Protocolo, otro grupo de expertos produciría las mismas recomendaciones y, en circunstancias clínicas semejantes, el Protocolo es interpretado y aplicado de la misma manera por distintos profesionales
Aplicabilidad Clínica	Los grupos de pacientes a los que es aplicable un Protocolo deben estar bien definidos, al nivel de especificación que permita la evidencia clínica y científica
Flexibilidad	Deben especificarse las excepciones conocidas y esperadas, en las que las recomendaciones no son aplicables
Claridad	El lenguaje utilizado no debe ser ambiguo, cada termino debe definirse con precisión, y deben utilizarse modos de presentación lógicos y fáciles de seguir.
Proceso Multidisciplinario	El proceso de elaboración de los protocolos debe incluir la participación de los grupos a quienes afecte. Esta participación puede consistir en formar parte de los paneles que desarrollan los protocolos, aportar evidencias y puntos de vista a estos paneles, y revisar los borradores de los protocolos.
Revisión explícitamente planificada	Los protocolos deben incluir información sobre cuándo deben ser revisados para determinar la introducción de modificaciones, según nuevas evidencias clínicas o cambios en los consensos profesionales
Documentación	Los procedimientos seguidos en el desarrollo de los Protocolos, los participantes implicados, la evidencia utilizada, las asunciones y razonamientos aceptados, y los métodos analíticos empleados, deben ser meticulosamente documentados y descritos.

Adaptado de: Field MJ; Lohr KN (eds.) *Clinical Practice Guidelines. Directions for a New Program.* Institute of Medicine National Academy Press. Washington DC; 1990

7. DISEÑO TOTAL DE LOS SERVICIOS: LA ÚLTIMA GENERACIÓN DE LOS PROTOCOLOS

El diseño conjunto de la actuación clínica y la forma de organizarse para realizarla es una actividad recientemente incorporada a los programas de gestión de calidad, en relación tanto con la idea de la protocolización como con el enfoque de reingeniería de procesos.

Sobre la idea de la protocolización, no ya sólo de la actuación clínica sino también de la forma de organizarse y actuar en equipo, junto a determinada situación o tipo de paciente, se han desarrollado en los últimos años una serie de metodologías para aplicación local, en un contexto concreto. Estos métodos y herramientas que intentan diseñar *todas* las actuaciones de *todo* el personal de un determinado proceso asistencial desde el principio al final reciben nombres diversos tales como "mapas de cuidados" (*care maps*), caminos críticos (*critical paths*), cuidados en colaboración (*colaborative care*), cuidados en cooperación (*cooperative care*) y protocolos de atención coordinada (*coordinated care*). Toda esta "última generación" de protocolos se ha desarrollado sobre todo en hospitales y va unida para algunos autores a la llamada *reingeniería de procesos*, un enfoque de diseño de la calidad con un gran predicamento en la actualidad. Hay, sin embargo, otros muchos métodos y enfoques nacidos y experimentados en la industria, los cuales están siendo adaptados y experimentados también en los servicios de salud.

8. MÉTODOS DE MEJORA DE LA CALIDAD CON BASE EN EL DISEÑO

La *reingeniería de procesos*, es hoy día una de las *buzzwords (palabras de moda)* en los ámbitos de la gestión de la calidad, pero no la única. En realidad, los métodos y conceptos relacionados con el diseño de la calidad son actualmente la punta de lanza en este campo. La Tabla 19.5 contiene una relación de los más prominentes, cuyo significado, a excepción de la protocolización ya comentada, resumimos muy brevemente a continuación.

TABLA 19.5. Métodos de mejora de la calidad con base en un enfoque de diseño.

La reingenieria de procesos, el benchmarking, el QFD, el método Taguchi y la planificación Hoshin son todos ellos métodos para la mejora de la calidad con un enfoque de diseño

- PROTOCOLIZACIÓN DE ACTIVIDADES (CLÍNICAS Y NO CLÍNICAS)
- REINGENIERÍA DE PROCESOS
- BENCHMARKING
- QFD (Quality Function Deployment, Despliegue de la función de calidad)
- MÉTODO TAGUCHI
- PLANIFICACIÓN HOSHIN

8.1. REINGENIERÍA DE PROCESOS

En su origen, se define como *reinventar* la manera de organizarse y hacer las cosas para conseguir saltos cualitativos sin precedentes. Descrito como un proceso altamente imaginativo (y arriesgado) parte, como todo diseño, de la definición de unos objetivos y resultados, para conseguir los cuales se *rediseña* la forma de hacer las cosas, en el convencimiento de que la forma actual (sin pararse necesariamente a evaluarlo) es obsoleta y/o inconsistente.

8.2. BENCHMARKING

Palabra sin traducción clara al castellano, significa el estudio y rediseño de los procesos para igualar y superar al centro o empresa que lo haga mejor. Esta "marca" ("benchmark") o estándar que ostenta el mejor en relación al servicio o aspecto del servicio que queremos mejorar es el punto de partida u objetivo en cuanto a resultados que hay que, como mínimo, igualar; y la forma en que lo consigue, un punto de partida como proceso a imitar. Previamente hay que definir, naturalmente, cuál es el servicio o aspecto del servicio que queremos someter a *benchmarking*; decisión en la que interviene no sólo nuestro propio nivel (que puede ser incluso bueno en términos de lo que sea normal en nuestro medio) sino también lo importante que sea para nuestra actividad en función de los servicios que ofrecemos y tipo de clientes/usuarios para los que trabajamos.

8.3. QFD

La herramienta básica del QFD es una matriz que relaciona los requisitos de calidad y las actividades del proceso con los que se asocian.

Es un método muy estructurado que tiene como herramienta básica una matriz que relaciona los requisitos de calidad a conseguir (identificados a partir de las necesidades y expectativas a cubrir por el servicio a diseñar) y las actividades con que se asocian. Adicionalmente, se cuantifican y representan los niveles de los que partimos para identificar a partir de ahí, en qué debemos incidir más en nuestro diseño para mejorar. Es probablemente el método que más claramente responde al esquema metodológico básico de diseño de la calidad. Puede incluir el enfoque de benchmarking a la hora de identificar en otros, procesos que consiguen lo que nosotros queremos conseguir.

8.4. MÉTODO TAGUCHI

El método Taguchi, a través de la "función de pérdidas" incorpora los costes de la calidad al diseño de los productos como un resultado a optimizar

El método Taguchi basa sus diseños en una definición previa de la calidad optima de los productos, servicios (o sus aspectos concretos), sobre la base de la llamada "función de pérdidas" que trata de establecer el punto en que se minimiza el coste de la mala calidad, dentro de la variabilidad posible de los niveles aceptables de calidad (llamados en la jerga industrial niveles de *tolerancia*) de los requisitos de calidad examinados. Adicionalmente, se averigua mediante experimentación, con métodos peculiares de Taguchi, el mejor diseño o correlación de factores que conduce a los niveles de calidad (y tolerancia) óptimos, tal como han sido previamente definidos. Uno de los grandes atractivos de este método es la incorporación explícita del coste de la calidad deficiente en la definición de calidad para la cual se realiza el diseño; aparte de la originalidad de los métodos de experimentación que propone.

8.5. PLANIFICACIÓN HOSHIN

Parte de la definición de los objetivos de calidad a conseguir, según los cuales se establece una planificación de actividades con responsabilidades explícitamente compartidas para toda la organización, cuyos componentes establecen en su nivel respectivo sus propios objetivos, en plena congruencia y complementariedad unos con otros. El método incluye asimismo, una estricta planificación

operativa, con revisiones como mínimo anuales, cuyos resultados pueden hacer que se redefinan los objetivos y el plan establecido.

A pesar de lo breve de la reseña que hemos efectuado, debe haber quedado claro por qué es el diseño de la calidad el componente de los programas de gestión de calidad más apasionante y el que más interés está despertando en la actualidad. También es, desafortunadamente, el más complicado. En este sentido, nadie debe sentirse desbordado por la retahíla de métodos y técnicas que podemos haber mencionado. Todo se puede ir analizando, comprendiendo y practicando si se afronta paso a paso.

BIBLIOGRAFÍA

- Horovitz J. La Calidad del servicio. Madrid: Mc Graw Hill; 1991.
- Field MJ, Lohr KN (editores). Clinical Practice Guidelines. Directions for a new program. Washington: Institute of Medicine. National Academy Press ; 1990.
- Vázquez JR, De León JM. Punto de partida conocimiento de necesidades y expectativas. En: Tratado de Calidad Asistencial. Tomo III. Madrid: Dupont Pharma; 1997. p. 15-38.
- González M. Urís J. Los grupos focales y su utilidad en el diseño de la Calidad. En: Tratado de Calidad Asistencial en Atención Primaria. Tomo III. Madrid: Dupont Pharma; 1997. p. 39-70.
- Saura J. Construcción y evaluación de protocolos o guías para la práctica. En: Tratado de Calidad Asistencial en Atención Primaria. Tomo III. Madrid: Dupont Pharma; 1997. p. 71-96.
- De la Puerta E. El despliegue de la función de Calidad (QFD). En: Tratado de Calidad Asistencial en Atención Primaria. Tomo III. Madrid: Dupont Pharma; 1997. p. 97-112.
- Parra P. Benchmarking. Diseñar los procesos imitando y superando al mejor. En: Tratado de Calidad Asistencial en Atención Primaria. Tomo III. Madrid: Dupont Pharma; 1997. p. 113-134.
- Udaondo M. Reingeniería. Evaluación y mejora integral de los servicios de salud. En: Tratado de Calidad Asistencial en Atención Primaria. Tomo III. Madrid: Dupont Pharma; 1997. p. 135-163.
- García JR. El método Taguchi, la planificación Hoshin y otros métodos de planificación de la Calidad en la industria. En: Tratado de Calidad Asistencial en Atención Primaria. Tomo III. Madrid: Dupont Pharma; 1997. p. 165-197.

FE DE ERRATAS

UNIDAD TEMÁTICA 5

- Página 64, Figura 5.4, donde dice *"Cambio por convenci-"*, debe decir *"Cambio por convencimiento"*
- Página 66, Figura 5.5, donde dice *"facilitadad"*, debe decir *"facilidad"*
- Página 66, Figura 5.5, donde aparece *"17"*, debe aparecer *"7"*

UNIDAD TEMÁTICA 7

- Página 100, Figura 7.2, las tareas sin símbolos deben enmarcarse en rectángulos
- Página 102, Figura 7.3, las tareas sin símbolos deben enmarcarse en rectángulos

UNIDAD TEMÁTICA 8

- Página 117, donde dice *"esfignomanómetro"*, debe decir *"esfigmomanómetro"*
- Página 122, párrafo cuarto, donde dice *"... si las consideraciones como cumplimiento ..."*, debe decir *"... si las consideramos como cumplimiento ..."*

UNIDAD TEMÁTICA 9

- Página 145, Tabla 9.1, falta destacar la fila correspondiente a 0,10
- Página 147, penúltimo párrafo, donde dice *"... cual es el método de que vamos a ..."*, debe decir *"... cual es el método que vamos a ..."*

UNIDAD TEMÁTICA 10

- Página 164, párrafo segundo, donde dice *"c^2"*, debe decir *"χ^2"*

UNIDAD TEMÁTICA 11

- Página 188, Figura 11.6, donde aparecen *"W"*, deben aparecer *"$*$"*
- Página 191, Tabla 11.5, donde dice *"... relativa y (porcentaje sobre el total de incumplimientos de la muestra) de incumplimientos, por cada criterio."*, debe decir *"... relativa (porcentaje sobre el total de incumplimientos de la muestra), por cada criterio."*

UNIDAD TEMÁTICA 12

- Página 201, Figura 12.1, donde aparece *"(c^2)"*, debe aparecer *"(χ^2)"*

UNIDAD TEMÁTICA 13

- Página 217, último párrafo, donde dice *"Tomemos el ejemplo el criterio ..."*, debe decir *"Tomemos el ejemplo del criterio ..."*
- Página 219, Tabla 13.2, faltan los asteriscos de la fila correspondiente a 0,23
- Página 223, penúltimo párrafo, donde dice *"... que más han mejorado con el 1 y el 3"*, debe decir *"... que más han mejorado son el 1 y el 3"*

- Página 224, en fórmula, donde dice *"parámetro estimado ± z * desviación estándar del parámetro estimado"*, debe decir *"parámetro estimado ± z•desviación estándar del parámetro estimado"*
- Página 228, sustituir la última fórmula por

$$z = \frac{0{,}98 - 0{,}9}{\sqrt{(0{,}94)\,(0{,}06)\,(\frac{1}{60} + \frac{1}{60})}} = 1{,}85$$

- Página 230, Tabla 13.6, columna de Mejora Relativa, fila Criterio 8, donde dice *"42,5%"*, debe decir *"28%"*
- Página 236, en la última fórmula, donde dice *"d_2?n_1"*, debe decir *"$d_2 \bullet n_1$"*
- Página 237, Figura 13.6, las tres gráficas deben ordenarse en 13.6a, 13.6b y 13.6c, consecutivamente

UNIDAD TEMÁTICA 15

- Página 266, último párrafo lateral, donde dice *"... riesgos a y b ..."*, debe decir *"... riesgos α y β ..."*

UNIDAD TEMÁTICA 16

- Página 278, Objetivo Específico 7, donde dice *"... entre los gráficos x, ε, p y u."*, debe decir *"... entre los gráficos x, $\bar{x}$, p y u."*
- Página 285, Figura 16.3, donde aparecen *"α"*, deben aparecer *"σ"*
- Página 291, Tabla 16.3, columna Tipo de Gráfico, donde dice *"Gráfico " " o de Medias ..."*, debe decir *"Gráfico "$\bar{x}$" o de Medias ..."*

UNIDAD TEMÁTICA 17

- Página 310, Tabla 17.4, en Cálculos para la línea promedio, se debe incluir *"LCI=9,3-(3•2.1/1,128)=4"*
- Página 312, último párrafo, donde dice *"... eleva el error a como mucho ..."*, debe decir *"... eleva el error α como mucho ..."*
- Página 315, segunda línea, eliminar símbolo *"(Σ)"*

UNIDAD TEMÁTICA 18

- Página 332, penúltimo párrafo, donde dice *"... ha que realizar concurrentemente ..."*, debe decir *"... hay que realizar concurrentemente ..."*
- Página 335, primer párrafo, donde dice *"... numero de eventos que mide al indicador ..."*, debe decir *"... numero de eventos que mide el indicador..."*
- Página 341, Paso 5, donde dice *"... z/2,58 ..."*, debe decir *"... $z \geq 2{,}58$..."*

Región de Murcia
Consejería de Sanidad
y Consumo

UNIVERSIDAD
DE MURCIA

MURCIA