



Serie
monografías
de apoyo a la
acreditación
N° 1

Indicadores de calidad: Manual de muestreo para prestadores

Intendencia de Prestadores

Agosto - 2010



Indicadores de calidad: Manual de muestreo

2010

Autor

Dr. Miguel Araujo

Intendencia de Prestadores

Colaboraron en la elaboración y edición del documento final:

Profesionales de la Intendencia de Prestadores:

- Dr. Rodrigo Contreras, Asesor Médico
- EU Carmen Monsalve, Departamento de Evaluación
- EU Ana María Cambón, Departamento de Evaluación
- EU Scarlett Morales, Departamento de Evaluación
- EU Benedicto Romero, Departamento de Evaluación

Por la Sociedad Chilena de Calidad Asistencial:

- TM Mariana Silva
- Dr. Juan Eduardo Sánchez
- Dra. Marcia Poblete

Otros temas de la serie:

- Indicadores de calidad: umbrales e interpretación de resultados
- Trazabilidad en laboratorio, patología y medicina transfusional
- Eventos adversos área médica
- Eventos adversos área quirúrgica
- Eventos adversos en atención abierta
- Eventos adversos en atención cerrada
- Eventos adversos por medicamentos
- Importancia de la formación, experiencia y volumen de atención de los profesionales
- Bioética: conceptos básicos para la acreditación de prestadores

Contenido

Indicadores de calidad:	1
Manual de muestreo para prestadores.....	1
Síntesis de Recomendaciones	5
I.Introducción	6
II.¿Porqué utilizar una muestra y no el total de casos?	6
III.¿Cuándo podría ser razonable medir el total de casos?	7
IV.¿Qué atributos debe tener la muestra? ¿Qué debemos entender por una muestra válida, representativa, confiable, o significativa?	7
1. Representatividad	7
a. Muestreo aleatorio	7
b. Muestreo sistemático.....	13
c. ¿Qué hacer si no es factible obtener una muestra probabilística?.....	13
2. Precisión y tamaño muestral.....	14
V.¿Cómo calcular el tamaño de muestra en la práctica?.....	17
1. Calculadora en Excel y programas disponible en Internet.....	17
2. Tablas	18
3. ¿Qué es un muestreo estratificado y cuándo se justifica hacerlo?.....	19
VI.Referencias.....	19
ANEXO: FÓRMULAS PARA EL CÁLCULO DE TAMAÑO MUESTRAL PARA INDICADORES DE CALIDAD	21

Síntesis de Recomendaciones

- Muchas veces los datos necesarios para construir un indicador no se encuentran disponibles en un sistema informático, lo que obliga a revisar fichas clínicas u otros registros individualmente. En tales casos, trabajar con muestras permite efectuar mediciones confiables y razonablemente precisas, ahorrando tiempo y recursos.
- La mejor forma de asegurar la representatividad de la muestra, es utilizar métodos probabilísticos de muestreo, tales como el muestreo aleatorio o sistemático. Con los programas disponibles hoy en día en Internet, seleccionar una muestra aleatoria resulta sencillo, siempre que podamos numerar secuencialmente los registros. Cuando el universo de fichas es grande, puede resultar más cómodo y simple hacer la selección en forma sistemática.
- Para estimar cuántos registros deberá analizar, utilice las fórmulas que le entregamos en **TamañoMuestra-Indicadores.xls** y calcule fácilmente el tamaño de muestra para indicadores basados en variables dicotómicas (porcentajes) y continuas (promedios).
- En los indicadores que se expresan como porcentajes, el parámetro más importante para estimar la muestra es la proporción de cumplimiento esperado del indicador, antes que el total de fichas o registros disponibles. Tanto en ese tipo de indicadores como en los referidos a variables continuas, el otro factor a considerar es el margen de error máximo aceptable, que suele fijarse entre un 5% y 10%.

I. Introducción

Una de las preguntas recurrentes entre quienes deben llevar a la práctica la construcción de indicadores o su medición, es la que se refiere a cuántos registros o fichas deben ser revisados para obtener una muestra adecuada en tamaño, y cómo seleccionarlos. Abordaremos aquí algunos conceptos clave que todo encargado de calidad y los profesionales ligados al tema deben conocer al respecto.

Nos referimos específicamente a los indicadores de datos agregados, es decir, aquellos que miden el desempeño basándose en procesos o eventos que ocurren con cierta frecuencia, y cuyos resultados generalmente se expresan en términos de:

- Porcentajes o proporciones, cuando la variable evaluada es dicotómica (cumple/no cumple). Por ejemplo, porcentaje de cirugías mayores con visita pre anestésica.
- Promedios, cuando la variable evaluada es continua. Por ejemplo, tiempo promedio de espera en pacientes categorizados en el servicio de urgencia.

Estos conceptos también son aplicables cuando se trata de obtener muestras para indicadores basados en la opinión de los pacientes a través de encuestas.

II. ¿Porqué utilizar una muestra y no el total de casos?

Porque normalmente no es necesario medir el total de casos para evaluar un proceso o un resultado de manera confiable. Revisar fichas o registros retrospectivamente, implica un trabajo arduo, y sus resultados deben compensar el esfuerzo. Los indicadores deben aportar datos que reflejen razonablemente la realidad, pero calcular un indicador no es una investigación científica. Los indicadores simplemente “INDICAN”, dan señales sobre el comportamiento del proceso o el resultado en la institución, su tendencia en el tiempo, y qué tan alejados nos encontramos de los niveles de cumplimiento que hemos definido como aceptables. Por lo tanto, no es un pecado que posean cierto margen de imprecisión o error.

El error que se comete debido al hecho de que se obtienen conclusiones sobre cierta realidad a partir de la observación de sólo una parte de ella, se denomina *error de muestreo*.

III. ¿Cuándo podría ser razonable medir el total de casos?

Cuando los datos son fácilmente accesibles y se pueden procesar sin necesidad de digitalarlos o codificarlos. Es decir, cuando están allí, “listos para ser usados”. El ejemplo más claro es cuando los datos con los que se construye el indicador están en un registro o base de datos informática, y lo que se requiere simplemente es resolver cómo analizarlos.

También cuando los casos (denominador) son pocos. Un muestreo en tales condiciones podría resultar en un indicador poco fiable.

IV. ¿Qué atributos debe tener la muestra? ¿Qué debemos entender por una muestra válida, representativa, confiable, o significativa?

Las muestras para calcular los indicadores deben cumplir dos condiciones esenciales:

1. Representatividad

La muestra debe reflejar, razonablemente:

- Lo acontecido a lo largo de todo el período que se pretende evaluar. Un error común es revisar registros consecutivos de un período reciente (por ejemplo, las fichas de los pacientes operados la semana anterior). Por lo tanto, si el indicador es semestral, la muestra debe extraerse de entre el total de casos ocurridos en el semestre, y no de una ventana de tiempo dentro de éste.
- La calidad promedio de las atenciones prestadas durante el período. La forma ideal de lograr esto es utilizar un método de muestreo probabilístico, aleatorio o sistemático:

a. Muestreo aleatorio

Para realizarlo es necesario disponer de un registro de los pacientes atendidos, que pueda ser numerado secuencialmente. Por ejemplo, imagine que su sistema informático le permite extraer una lista de las cirugías mayores realizadas en el semestre pasado en formato Excel¹:

¹ Los datos de la tabla son ficticios.

RUT	N°FICHA	SEXO	EDAD	CIRUGIA
1643298-9	8541	M	18	APENDICECTOMIA
3532116-5	354	F	62	COLECISTECTOMÍA
6551687-4	3218	F	45	HISTERECTOMIA
8735213-9	3658	M	36	PLEUROTOMIA MINIMA
14629872-2	2874	M	18	APENDICECTOMIA
9875123-8	669	M	74	HERNIOPLASTÍA
3548421-1	2185	F	55	COLECISTECTOMÍA
5420101-6	742	M	45	COLECISTECTOMÍA
11542913-0	6821	F	46	SAFENECTOMIA
7635094-9	4555	M	29	OSTEOSINTESIS

Lo que debe hacer a continuación es numerar secuencialmente la lista:

#	RUT	N°FICHA	SEXO	EDAD	CIRUGIA
1	1643298-9	8541	M	18	APENDICECTOMIA
2	3532116-5	354	F	62	COLECISTECTOMÍA
3	6551687-4	3218	F	45	HISTERECTOMIA
4	8735213-9	3658	M	36	PLEUROTOMIA MINIMA
5	14629872-2	2874	M	18	APENDICECTOMIA
6	9875123-8	669	M	74	HERNIOPLASTÍA
7	3548421-1	2185	F	55	COLECISTECTOMÍA
8	5420101-6	742	M	45	COLECISTECTOMÍA
9	11542913-0	6821	F	46	SAFENECTOMIA
10	7635094-9	4555	M	29	OSTEOSINTESIS

Sabiendo cuánta es la muestra necesaria (ver punto b.), puede ahora seleccionar una muestra al azar de esas 10 fichas. Para ello puede ocupar cualquiera de los diversos programas disponibles en Internet (aquí utilizaremos 2 como ejemplo pero hay otros).

Ejemplo 1

La página *Stat Trek* ofrece una calculadora simple para seleccionar muestras aleatorias en <http://stattrek.com/Tables/Random.aspx>.

Imagine que debe seleccionar aleatoriamente 5 fichas de las 10 de la lista. Ingrese los datos en la calculadora:

Note: The seed value is optional. Leave it blank to generate a new set of numbers. Use it to repeat a previously-generated set of numbers.

How many random numbers?

Minimum value

Maximum value

Allow duplicate entries

Seed (optional)

Lo que hemos dicho al programa es que seleccione 5 fichas, de entre una lista que comienza en 1 y termina en 10, sin duplicados (sin que pueda aparecer seleccionada una misma ficha más de una vez) para lo cual debemos seleccionar la opción “False”. Deje “Seed” en blanco. Al oprimir el botón “Calculate” obtendremos las 5 fichas que deberemos extraer para el análisis (3, 2, 7, 8 y 4):

5 Random Numbers*


0 3 0 2 0 7 0 8 0 4

* This table of 5 random numbers was produced according to the following specifications: Numbers were randomly selected from within the range of 1 to 10. Duplicate numbers were not allowed.

Ejemplo 2

El sitio *Research Randomizer* provee también una funcionalidad similar en <http://www.randomizer.org/form.htm>

Seleccionaremos esta vez una muestra aleatoria de 4 fichas de entre las 10 de la lista. Ingrese los datos en la calculadora:



To generate random numbers, enter your choices below (using integer values only):

How many sets of numbers do you want to generate? [Help](#)

How many numbers per set? [Help](#)

Number range (e.g., 1-50): From: To: [Help](#)

Do you wish each number in a set to remain unique? [Help](#)

Do you wish to sort the numbers that are generated? [Help](#)

How do you wish to view your random numbers? [Help](#)

Lo que hemos dicho al programa es que seleccione 1 set de 4 fichas, de entre una lista que comienza en 1 y termina en 10, sin duplicados, y que los ordene de menor a mayor. Al oprimir el botón “Randomize Now!” obtendremos las 4 fichas que deberemos extraer para el análisis (2, 5, 8, 9):

Research Randomizer Results

1 Set of 4 Unique Numbers Per Set
Range: From 1 to 10 -- Sorted from Least to Greatest


Job Status: Finished

Set #1:

2, 5, 8, 9

Este programa nos permite también utilizar algunas opciones de muestreo más complejas, pero necesarias en un momento dado. Por ejemplo, imagine que desea aplicar un indicador que mida la utilización de medidas de prevención de caídas en los pacientes hospitalizados del servicio de medicina. Se decide obtener resultados semestrales, y el tamaño de muestra calculado fue 120, lo que implica 20 pacientes al mes. Para la enfermera que será responsable de la medición, lo más práctico es ir aplicando todas las semanas la pauta de supervisión a un grupo de pacientes, digamos 5 casos semanales. ¿Cómo podría planificarse un muestreo aleatorio en este caso?:

1. Seleccione en primer lugar el día de la semana en que se realizará la supervisión. Ello puede programarse de antemano para todo el semestre. Cada semestre tiene en total 26 semanas, necesitamos entonces 26 “sets” de números aleatorios. Como seleccionaremos un día de cada semana, cada set será un solo número, entre 1 y 7, que corresponderá a uno de los 7 días de la semana. En el formulario esto se traduce en las siguientes entradas:



To generate random numbers, enter your choices below (using integer values only):

How many sets of numbers do you want to generate? [Help](#)

How many numbers per set? [Help](#)

Number range (e.g., 1-50):

From:

To:

Nota: Si desea seleccionar sólo días hábiles (lunes a viernes), bastará definir que el rango estará entre 1 y 5, en vez de 1 y 7.

- Al oprimir el botón “Randomize Now!” obtendremos, para cada semana, el día en que se deberá realizar la supervisión:

Research Randomizer Results

26 Sets of 1 Unique Numbers Per Set
Range: From 1 to 7 -- Unsorted

Job Status: **Finished**

Set #1:

3

Set #2:

4

Set #3:

2

Set #4:

7

- Por lo tanto la semana 1 (Set #1) la supervisión correspondería hacerla el día 3 (miércoles), la semana 2 (Set #2) el día 4 (jueves), la semana 3 (Set #3) el día 2 (martes), la semana 4 (Set #4) el día 7 (domingo), y así sucesivamente.
- Ahora necesitamos saber cuáles serán los pacientes a los que se aplicará la pauta de supervisión. Hemos dicho que analizaremos 5 pacientes cada vez. Si el servicio posee por ejemplo 40 camas, necesitamos seleccionar, para cada una de las 26 semanas, 5 camas en forma aleatoria. En este caso la entrada de datos sería:



To generate random numbers, enter your choices below (using integer values only):

How many sets of numbers do you want to generate? [Help](#)

How many numbers per set? [Help](#)

Number range (e.g., 1-50): From: To: [Help](#)

Do you wish each number in a set to remain unique? [Help](#)

Do you wish to sort the numbers that are generated? [Help](#)

Donde: el número de sets es 26, que corresponde a las semanas del semestre; 5 es el número de camas que se evaluará cada semana; 1-40 son las camas del servicio, que van desde la 1 a la 40. Pedimos además al programa que ordene los números de cada set de menor a mayor (Least to Greatest).

- Al oprimir el botón “Randomize Now!” obtendremos ahora, para cada semana, las camas en las que se deberá realizar la supervisión:

Research Randomizer Results

26 Sets of 5 Unique Numbers Per Set
Range: From 1 to 40 -- Sorted from Least to Greatest

Job Status: **Finished**

Set #1:

3, 14, 15, 33, 37

Set #2:

1, 15, 18, 32, 38

Set #3:

18, 20, 23, 28, 30

Set #4:

5, 11, 12, 16, 23

6. Por lo tanto, la semana 1 (Set #1), cuyo día previamente seleccionado es el miércoles, la supervisión debiera hacerse en las camas 3, 14, 15, 33 y 37, la semana 2 (Set #2), cuyo día previamente seleccionado es el jueves, la supervisión debiera hacerse en las camas 1, 15, 18, 32, 38, y así sucesivamente.

¿Complicado? No en realidad, la programación puede hacerse para todo el año por anticipado, y podremos trabajar con una muestra confiable, sin necesidad de supervisar a todos los pacientes cada vez.

b. Muestreo sistemático

Consiste en extraer los casos a un intervalo regular predefinido. El intervalo se obtiene dividiendo el total de casos, fichas o pacientes, por la muestra. Por ejemplo, si el total de operaciones fueron 2.500, y deseamos seleccionar 100, deberemos seleccionar una de cada 25 ($2.500 / 100 = 25$).

La ventaja de este método de muestreo es que no se requiere numerar secuencialmente las fichas o registros, como exigía el muestreo aleatorio, y por lo tanto es fácilmente aplicable cuando los registros se encuentran, por ejemplo, almacenados en papel en una carpeta o archivador. También es más fácil de utilizar cuando, pese a existir un registro computacional, el número total de casos es muy alto.

No es necesario que la primera ficha seleccionada sea a la vez la primera del total de fichas disponibles, puede ser la tercera, la quinta, etc. De hecho, y siguiendo el ejemplo, idealmente debiera ser seleccionada también al azar de entre las primeras 25.

NOTA: Cualquiera sea el método de muestreo utilizado, antes de iniciar la obtención de casos para una muestra, debe definirse el mecanismo de sustitución, es decir, cómo vamos a reemplazar aquellos casos que, seleccionados con un determinado método, no cumplen con los criterios de inclusión definidos previamente o no contienen datos válidos. Por ejemplo, podemos tomar el caso inmediatamente siguiente o anterior. El método de sustitución debe ser siempre el mismo en todo el muestreo.

c. ¿Qué hacer si no es factible obtener una muestra probabilística?

Los métodos de muestreo no probabilístico (cualquiera distinto al muestreo aleatorio o sistemático descritos en la sección previa) no garantizan la representatividad de la muestra, y por lo tanto no permiten inferir que los resultados del indicador reflejan fielmente la realidad de lo que se quiere medir.

En ocasiones, no obstante, existen limitaciones prácticas que hacen difícil realizar un muestreo probabilístico en regla, y es común que los indicadores se construyan a partir de otras formas de muestreo, que denominaremos en general “por conveniencia” (una forma de decir que se adaptan a las rutinas o facilidades de quien debe recoger los datos, lo que no implica que sean “convenientes” desde el punto de vista metodológico). Cuando quien realiza el muestreo lo hace seleccionando lo que le parece una muestra representativa, de acuerdo a su apreciación subjetiva, se habla de muestreo “por juicio”.

Un ejemplo de este tipo de muestreo sería aquél en que, para verificar el cumplimiento de las medidas de prevención de escaras definidas en el protocolo institucional, la enfermera coordinadora decidiera aplicar discrecionalmente una vez por semana una pauta de supervisión a algunos de los pacientes hospitalizados ese día, y que el día en cuestión lo decidiera en función de la carga de trabajo, cuando tuviera un tiempo disponible.

Asumiendo que este método tiene un alto potencial de entregar una muestra sesgada, quienes lo apliquen deben procurar incorporar el máximo de azar en la selección, por ejemplo, que el día en que se realice la supervisión no sea conocido previamente por el personal, que la fecha se determine lo más aleatoriamente posible, que se incluya todos los pacientes en la supervisión o una muestra sistemática de ellos, y que quien audita intente evitar cualquier sesgo de selección evidente (por ejemplo, esperar un día en que las condiciones sean más “propicias” para la supervisión, porque hay personal con mejor actitud hacia la prevención).

Pese a estos resguardos, la interpretación de los resultados de un indicador así calculado debiera realizarse con precaución, y la institución debiera avanzar hacia la implementación de indicadores basados en algún método de muestreo probabilístico.

¿Por qué este tipo de muestreo es propenso a seleccionar muestras sesgadas?

Porque existen procesos inconscientes que orientan las conductas de la gente hacia la obtención de resultados “favorables”, y al menor gasto de energía en el proceso. Como la construcción de los indicadores en los servicios clínicos es realizada en general por personal del propio servicio, existirá la tendencia a demostrar que los procesos se cumplen de manera adecuada, antes que a medir objetivamente el desempeño. Y también se tenderá a examinar los casos más fácilmente accesibles (por ejemplo, los más cercanos a la central de enfermería, que son a la vez los que probablemente reciben mejores cuidados) y en los días en que resulte más cómodo hacerlo.

2. Precisión y tamaño muestral

Asumiendo que la muestra es representativa, el resultado del indicador tendrá de todas maneras un margen de error por azar que debe ser considerado. Para que este margen no sea demasiado amplio, es importante que la muestra tenga el tamaño adecuado (ley de los grandes números). Al aumentar la muestra el potencial de error se hace menor y viceversa. Decimos que la precisión es poca si el potencial de error es amplio, y alta si es pequeño. Los intervalos de confianza (IC) nos dan alguna idea de la precisión de un

indicador. Por convención y familiaridad, se usan IC 95% en la mayoría de los casos y eso no vale la pena variarlo.

La precisión hay que considerarla al evaluar las diferencias entre una medición y otra. Si un indicador arrojó un 78% de cumplimiento con un IC95% 63% a 92%, eso implica que el nivel de cumplimiento real en la población de la que se extrajo la muestra podría estar, con un 95% de confianza o seguridad, entre 63% y 92%. Si una nueva medición entrega 73% con un IC95% 61% a 85%, ambos IC se superponen, y por lo tanto los resultados no son estrictamente distintos, aunque el segundo resultado esté 5 puntos bajo el primero

El tamaño muestral dependerá principalmente de la proporción esperada de cumplimiento del proceso que se está midiendo. Mientras más se aleje el resultado esperado del indicador, del 50% de cumplimiento (en un sentido u otro, hacia el 0% o hacia el 100%), menor será la muestra necesaria. Así, un cumplimiento esperado de 20% requerirá una muestra menor que uno de 40%. Una proporción esperada de 50% se considera “el peor de los escenarios”, es decir, aquel valor esperado que requiere una muestra con el mayor número de casos. Es importante insistir en esto, porque existe la concepción errónea de que la muestra depende sobre todo de cuánta es la población total (por ejemplo el total de pacientes hospitalizados en el período), y que debe representar un cierto porcentaje de ésta (por ejemplo, que deberían ser al menos un 5% o 10% del total de fichas). Si bien es cierto que cuando el número total de casos es relativamente pequeño (como suele ocurrir en este tipo de indicadores), el tamaño de la población total influye sobre el tamaño de la muestra, este último debe calcularse caso a caso, y no pensarlo en términos de porcentaje. Cuando el número total de casos es grande, el tamaño de muestra es independiente de éste (por ejemplo una muestra de 250 sujetos, sirve igual para evaluar una población de 5.000 o de 100.000 personas).

Las tablas siguientes ilustran el efecto de estos 2 parámetros –total de casos y frecuencia esperada de cumplimiento- sobre la muestra:

Aquí se observa cómo para un mismo porcentaje de cumplimiento de 90%, el tamaño muestral requerido varía apenas entre 82 y 131 para un total de casos entre 200 y 2.500, mientras que en términos porcentuales el rango de variación es muy amplio (lo cual demuestra que no se debe aplicar un porcentaje fijo como criterio para calcular el tamaño de muestra).

Total de casos	Frecuencia estimada de cumplimiento	Muestra estimada*	Porcentaje del total de casos
200	90%	82	41%
1.000	90%	122	12%
2.500	90%	131	5%

* Para un +/- 5% de error muestral

Como sea, cuando el total de casos es pequeño la muestra requerida tiende a ser mayor porcentualmente que cuando los casos son muchos. Si comparamos ahora los valores de la tabla con los calculados para un 50% de cumplimiento, podemos ver cómo la muestra estimada crece considerablemente.

Total de casos	Frecuencia estimada de cumplimiento	Muestra estimada*	Porcentaje del total de casos
200	50%	132	66%
1.000	50%	278	28%
2.500	50%	333	13%

* Para un +/- 5% de error muestral

Recuerde además que la muestra necesaria para un cumplimiento de 90% será la misma que para un 10%, y la calculada para un 30% será igual a la de un 70%, es decir, lo importante es cuánto se aleja el porcentaje del 50% (entre 90% y 50% existen 40 puntos porcentuales de diferencia, lo mismo que entre 10% y 50%).

Por último, el tercer factor que incide en forma importante en el tamaño de la muestra a utilizar, es el porcentaje de error que estamos dispuestos a asumir en el resultado. Los valores de la tabla anterior fueron calculados para un margen de error de +/- 5%. Eso implica que si la medición del indicador arrojó finalmente un 54% de cumplimiento, el valor real –si se analizara el total de casos- podría estar entre 49% y 59%.

Veamos qué ocurre cuando ampliamos ese margen a +/- 10%, es decir, si sacrificamos precisión.

Total de casos	Frecuencia estimada de cumplimiento	Muestra estimada*	Porcentaje del total de casos
200	50%	65	33%
1.000	50%	88	9%
2.500	50%	93	4%

* Para un +/- 10% de error muestral

En este caso, si bien podemos trabajar con muestras bastante menores, al interpretar ese 54% de cumplimiento debemos considerar que el valor verdadero podría estar en un rango mucho más amplio, entre 44% y 64%.

Por lo tanto, si decidimos trabajar con una muestra de casos, tendremos que estimar y tomar decisiones respecto a 2 supuestos básicos:

- Cuál es el porcentaje de cumplimiento que creemos va a arrojar el indicador: para ello podemos utilizar mediciones previas realizadas en el establecimiento, pero si no disponemos de ellas, lo recomendado es usar el escenario más conservador –el que exige mayor muestra-, que es de 50%.
- Con qué margen de error es razonable trabajar: en general este porcentaje variará entre 5% y 10%. Porcentajes mayores entregarán resultados demasiado imprecisos, que será difícil interpretar para quien debe tomar decisiones a partir de ellos.

Pese a todos los cálculos que podamos realizar, hay que considerar también los recursos humanos disponibles y el grado de esfuerzo que significa la medición. El desafío es lograr un adecuado balance entre dicho esfuerzo y la utilidad práctica que se dará a los

resultados del indicador en el proceso de mejoría de calidad. Ello implica que en ocasiones, por razones pragmáticas, pudiera ser necesario sacrificar rigor metodológico para hacer viable la aplicación del indicador. Hasta dónde es posible simplificar la evaluación de un proceso o resultado, manteniendo el sentido original de la medición y que ésta sea útil, corresponde a una decisión que los equipos locales deben hacer en cada caso.

V. ¿Cómo calcular el tamaño de muestra en la práctica?

El cálculo propiamente tal lo podemos hacer con cualquiera de los software estadísticos y calculadoras disponibles en forma gratuita en la web, con la ayuda de las fórmulas que hemos preparado en formato Excel (adjuntas a esta monografía), o bien apoyándonos en tablas de valores de tamaño muestral. No obstante, antes de usar estos recursos es recomendable conocer las fórmulas en que están basados, para entender cuáles son los datos que debemos considerar para hacer el cálculo, y que se nos pedirá ingresar (ver Anexo al final de esta Monografía).

1. Calculadora en Excel y programas disponible en Internet

Existen diversas páginas de Internet donde uno puede realizar estos cálculos, por ejemplo:

Para proporciones

- http://www.dimensionresearch.com/resources/calculators/sample_size.html
- http://department.obg.cuhk.edu.hk/ResearchSupport/Sample_Size_EstPrev.asp
- <http://www.raosoft.com/samplesize.html>
- <http://www.openepi.com/SampleSize/SSPropor.htm>
- http://www.macorr.com/ss_calculator.htm

Para promedios:

- http://department.obg.cuhk.edu.hk/researchsupport/Sample_size_EstMean.asp
- <http://www.aboriginemundi.com/index.php/2010/04/sample-size-calculator/>

O bien, lo más simple de todo, utilice las fórmulas en Excel adjuntas (TamañoMuestra-Indicadores.xls).

2. Tablas

2.1 Indicadores basados en proporciones

La Tabla siguiente permite estimar la muestra, según la proporción esperada de cumplimiento y la precisión deseada (amplitud total del intervalo de confianza):

Proporción esperada	Amplitud Total Intervalo de Confianza		
	0.10	0.20	0.30
0.10	139	-	-
0.15	196	-	-
0.20	246	62	-
0.25	289	73	-
0.30	323	81	36
0.40	369	93	41
0.50	384	96	43

Así, si la proporción esperada de cumplimiento es de 25% (o su recíproco, 75%), para un margen de error de +/-10% (amplitud total 0,20), la muestra necesaria serán 73 sujetos.

2.2 Indicadores basados en variables continuas

La Tabla siguiente permite estimar la muestra, según la DE de la variable, y el margen total de error deseado (en la unidad de medida original):

W/S	Intervalo de Confianza		
	90%	95%	99%
0.10	1.083	1.537	2.665
0.15	482	683	1180
0.20	271	385	664
0.25	174	246	425
0.30	121	171	295
0.35	89	126	217
0.40	68	97	166
0.50	44	62	107
0.60	31	43	74
0.70	23	32	55
0.80	17	25	42
0.90	14	19	33
1.00	11	16	27

W/S: Amplitud estandarizada del Intervalo de Confianza
 W: Error total deseado (en la unidad de medida original)
 S: Desviación estándar de la variable

Así, si se desea un error total no mayor de 30 (+/- 15) –por ejemplo minutos-, y la DE es 13 minutos, ello arroja un $W/S=2,3$ lo que implica una muestra aproximada de 246 sujetos (para un IC de 95%, que es el habitualmente considerado).

3. ¿Qué es un muestreo estratificado y cuándo se justifica hacerlo?

Consiste en seleccionar submuestras, procurando que se encuentren bien representadas dos o más categorías de sujetos que nos interesa analizar por separado o comparar entre sí, sean grupos de pacientes, áreas, servicios clínicos, períodos de tiempo, turnos, o miembros individuales del personal. Puede ser proporcional (manteniendo la relación de tamaño de los estratos en la población de referencia), o bien optimizado, cuando extraemos más sujetos de un estrato que los que proporcionalmente le corresponderían, para que queden mejor representados (por ejemplo, si la población de ese estrato es muy pequeña o se espera que tenga mayor variabilidad).

Dado que cada submuestra debiera ser autosuficiente en tamaño, y todas deben entregar resultados con un nivel de precisión razonable, para indicadores como los que aquí nos ocupan, es recomendable no abusar de la estratificación, porque obliga a aumentar la muestra global para que el resultado de cada estrato sea suficientemente representativo y preciso.

Por ejemplo, imagine que desea evaluar el cumplimiento de los criterios de transfusión del establecimiento en términos globales, pero también quiere analizar las solicitudes urgentes y electivas por separado. Supongamos que el 80% de las transfusiones son de urgencia y un 20% electivas. En términos globales, si Usted espera alrededor de un 90% de cumplimiento y asume un 10% de error, con una muestra de 35 sería suficiente. Proporcionalmente, la muestra de solicitudes electivas sería solamente de 7 (20% de 35), lo que claramente es insuficiente. Por ello, lo razonable sería extraer –en forma aleatoria o sistemática- 35 solicitudes electivas y otras 35 de urgencia (asumiendo que en ambos estratos el nivel de cumplimiento será similar).

Hecho esto, lo importante es que al calcular el resultado general se pondere los resultados de ambos estratos según las proporciones originales. Si el porcentaje de cumplimiento en urgencias fue de 75% y en electivas de 60%, el global será no el promedio aritmético (67,5%), sino el ponderado: $+ (0,75 \times 0,8) + (0,60 \times 0,2) = 0,60 + 0,12 = 0,72$ (72%).

VI. Referencias

1. The Joint Commission. 'Specifications Manual for National Hospital Inpatient Quality Measures', version 3.1a (applicable 4/1/2010 to 9/30/2010)

2. Jane Moore, Mandy Smith, Mary Barwick. Healthcare Quality Improvement Partnership (HQIP). An Introduction to Statistics for Clinical Audit
3. Healthcare Quality Improvement Partnership (HQIP). Criteria and indicators of best practice in clinical audit. 2009 www.hqip.org.uk
4. National Institute for Clinical Excellence. Principles for best practice in clinical audit. Oxford:Radcliffe Medical Press Ltd. NICE; 2002.

ANEXO: FÓRMULAS PARA EL CÁLCULO DE TAMAÑO MUESTRAL PARA INDICADORES DE CALIDAD

1. Indicadores basados en proporciones

En estos casos el cálculo dependerá de:

- El nivel de cumplimiento esperado para el indicador (p). Si podemos estimar cuál será el grado de cumplimiento local del proceso o resultado que se quiere medir, usaremos esa proporción para calcular la muestra. Si no tenemos noción sobre ello, lo recomendable sería entonces usar como supuesto un 50% ($p=0,5$), que es el que dará lugar a la muestra más alta.
- El margen de error máximo aceptable para nuestra medición (i), que en el caso de los indicadores habitualmente se fija entre un 5% y 10%.
- La población total o universo (pacientes, fichas, registros) de la cual pretendemos extraer la muestra (N):
 - Si la población no es conocida (escenario poco frecuente en nuestro caso, generalmente existe una noción clara de cuántos son los casos totales), la fórmula es:

$$n = Z_{\alpha}^2 \frac{p \cdot q}{i^2}$$

- Si el tamaño es conocido o se puede estimar (n° total de casos atendidos en el período, de cirugías realizadas, de pacientes hospitalizados, etc.) la fórmula es:

$$n = Z_{\alpha}^2 \frac{N \cdot p \cdot q}{i^2 (N - 1) + Z_{\alpha}^2 \cdot p \cdot q}$$

Donde:

- n** Tamaño de muestra
- N** Tamaño de la población
- Z** Valor estadístico de la curva normal de frecuencias: 1,96 para $\alpha = 0,05$ que es el habitualmente usado
- p** Proporción esperada de cumplimiento.
- q** $1-p$ (Si $p=0,25$, $q=0,75$)
- i** Error máximo esperado. Para un error del 10% ($i=0,1$), el resultado del indicador tendrá, con una seguridad del 95% (para $\alpha = 0,05$), un valor real posible 10% sobre o bajo el resultado puntual obtenido. Ello implica que si el resultado

puntual fue 76% de cumplimiento, el resultado en la población total podría estar entre 66% y 86%.

Ejemplo

Imagine que desea medir la proporción de cumplimiento de un procedimiento de enfermería en pacientes hospitalizados. La medición es trimestral y, de acuerdo a las estadísticas del establecimiento, usted calcula que en el período habrá un total estimado de 1.200 admisiones. Supervisiones previas sugieren que el nivel de cumplimiento del procedimiento se encontrará en torno al 60%. Usted desea que el indicador posea un error máximo de +/- 5%.

Reemplazando los valores en la fórmula obtendremos:

$$n = 1,96 * 1,96 \frac{1.200 * 0,60 * 0,40}{(0,05*0,05*1.199)+(1,96*1,96*0,60*0,40)} = 282$$

La muestra requerida es de 282 pacientes.

2. Indicadores basados en variables continuas

En este tipo de indicadores, cuyos resultados se expresan como promedios, se requiere una estimación de la desviación estándar del parámetro en la población, y el margen de error a considerar:

$$n = \left(z^* \frac{\sigma}{m} \right)^2$$

Donde:

- n** Tamaño de muestra
- Z** Valor estadístico de la curva normal de frecuencias: 1,96 para $\alpha = 0,05$ que es el habitualmente usado
- σ** Desviación estándar (DE).
- m** Margen de error esperado sobre o bajo el resultado puntual que se obtenga (en la unidad de medida original).

Ejemplo

Imagine que desea medir el tiempo promedio de espera de los pacientes en el Servicio de Urgencia. Basado en una muestra piloto de 50 atenciones, Usted primero calculó que la DE aproximada de esta variable es 16. ¿Qué refleja este valor de DE? Refleja el grado de

variación que tienen los tiempos de espera de los pacientes. Mientras mayor la DE, mayor será también el tamaño de muestra requerido. ¿Cómo calcular la DE? Traslade los valores medidos a una planilla Excel y use la función DESVEST para obtener el dato, ejemplo:

	A	B	C
1	Paciente	t Espera	
2	1	23	
3	2	25	
4	3	65	
5	4	11	
6	5	45	
7	6	35	
8	7	47	
9	8	55	
10	9	43	
11	
12	50	41	
13	PROMEDIO	39,0	
14	DE	16,0	

NOTA: Si el indicador ya fue calculado previamente, puede usar esos datos para obtener le DE estimada.

Enseguida debe definir cuál será el margen de error con que va a trabajar, expresada en la medida original, es decir, en minutos. Asumamos que fija ese error en +/- 5 minutos.

Reemplazando los valores en la fórmula obtendremos:

$$n = \left(1,96 * \frac{16}{5} \right)^2 = 6,27 * 6,27 = 40$$

La muestra requerida es de 40 pacientes.