

VALIDACIÓN DE MÉTODOS ANALÍTICOS CUALITATIVOS

Itziar Ruisánchez, Esther Trullols y F. Xavier Rius .

Grupo de Quimiometría y Cualimetría. Departament de Química Analítica i Química Orgànica.

Universitat Rovira i Virgili. Pl. Imperial Tàrraco, 1, 43005 Tarragona.

Es bien conocida la necesidad de validar los métodos de análisis y en concreto en este artículo se aborda la validación de los métodos de análisis cualitativos. Si bien el concepto de validación no depende de que éste sea cuantitativo o cualitativo, en la práctica los parámetros de calidad que caracterizan a los métodos cualitativos son distintos. Así, además de describir las características de los métodos cualitativos, se definen los principales parámetros de calidad como son: falsos positivos y negativos, sensibilidad, especificidad, región de incertidumbre, límite de corte, etc. Finalmente, se describen muy brevemente cuatro vías para establecer dichos parámetros.

Introducción a los sistemas de medida cualitativos

En los laboratorios de análisis es cada vez más frecuente la introducción de sistemas de medida de respuesta rápida que generan respuestas más de tipo cualitativo que cuantitativo. La respuesta cualitativa suele ser binaria del tipo SI/NO y puede responder a distintas situaciones: presencia/ausencia de un determinado analito en una muestra, presencia/ausencia por encima de un determinado nivel (normalmente de concentración) que habitualmente viene fijado por la legislación o el cliente, etc.

Estos sistemas se denominan habitualmente sistemas de *screening* o de cribado. Desde el punto de vista práctico, el principal interés en el desarrollo de estos sistemas radica en que se utilizan como una etapa previa de cribado de las muestras (figura 1), de manera que se evita así que todas las muestras sean sometidas a todo el proceso de medida químico. Únicamente seguirán el proceso de análisis cuantitativo aquellas muestras cuya respuesta sea un ‘SI’ (positivo), aquellas en las que se detecte la presencia de un analito o se detecte que está por encima del nivel permitido.

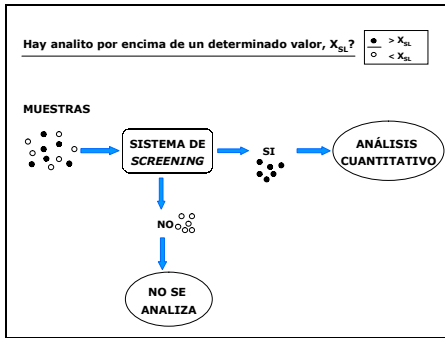


Figura 1. Sistemas de cribado (*screening*).

Tipos de sistemas de medida cualitativos.

Hacer clasificaciones siempre resulta difícil ya que depende de los criterios que se utilicen. Si atendemos al tipo de sistema utilizado para la obtención de la respuesta (detección) podemos diferenciar dos grandes grupos: análisis cualitativo clásico o sensorial y análisis cualitativo instrumental (Tabla 1).

<p>➤ Detección sensorial</p> <ul style="list-style-type: none"> □ Color: cambio, aparición, etc. □ Olor □ ELISA <p>➤ Detección instrumental</p> <ul style="list-style-type: none"> □ UV-Vis, Fluorescencia, Potenciometría, etc. □ Sensores: químicos, bio- químicos, etc. □ ELISA
--

Tabla 1. Tipos de sistemas de utilizados en el cribado. En el análisis cualitativo clásico la detección es de tipo sensorial y se realiza en base a los sentidos humanos como pueden ser la vista, olfato, etc. El más utilizado es la vista y la mayoría de sistemas se basan en la aparición o no de un determinado color como resultado

de una reacción química, bioquímica o inmunológica.

En este tipo de sistemas, la respuesta binaria ‘SI/NO’ se obtiene de forma directa, sin ningún tratamiento de los datos. Generalmente, la presencia de un analito se determina por comparación respecto a un blanco (muestra sin el analito), por ejemplo si se obtiene un determinado color hay analito, mientras que si no se obtiene (color del blanco) no lo hay.

Dentro de este grupo de sistemas cualitativos se encuentran los denominados ‘*tests kits*’. Son dispositivos comerciales diseñados para una aplicación concreta que contienen todos los reactivos necesarios y en algunos casos incluyen un sistema instrumental sencillo necesario para la obtención de la respuesta.

El otro gran grupo es el que hemos denominado análisis cualitativo instrumental. En este caso, la detección se realiza en base a una medida instrumental (colorimetría, fluorescencia, voltamperometría, etc.). La presencia o no de un determinado analito depende del nivel al que interese detectarlo. Bien puede ser al nivel del límite de detección o a un nivel superior que

generalmente corresponde al fijado por la legislación.

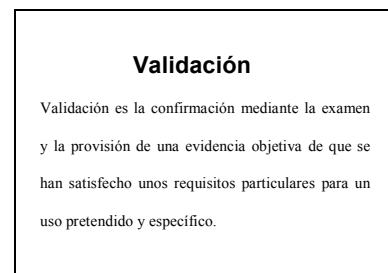
En este tipo de sistemas de medida, la respuesta instrumental que se obtiene debe transformarse en respuesta binaria del tipo SI/NO y por lo tanto implica un tratamiento de datos. Primero hay que establecer la respuesta instrumental, por ejemplo un valor de absorbancia para la concentración correspondiente al valor al que se quiere cribar (nivel de concentración establecido por la legislación) y debe compararse con la respuesta instrumental obtenida para cada muestra. Si es superior, hay analito por lo que la respuesta binaria es SI, y si el valor es inferior la respuesta binaria es NO.

Otro grupo de sistemas cualitativos que podemos considerar aparte son los que utilizan el método ELISA. Están basados en una reacción inmunológica y al igual que los *test kits*, la detección puede ser tanto sensorial como instrumental.

Validación de los sistemas de medida cualitativos

Al igual que los métodos de medida cuantitativos, los métodos cualitativos también deben validarse. La validación de un método no depende de que éste sea

cuantitativo o cualitativo, ya que validar consiste en verificar y documentar su validez, su adecuación a unos requisitos previamente establecidos. Esta es la idea que queda recogida en la definición proporcionada por la norma ISO 8402 (ver cuadro 1) y que implica el concepto de adecuación a la finalidad o propósito perseguido, '*fit-for-purpose*'.



Cuadro 1, Definición de validación según la norma ISO 8402

Por lo tanto, ante cada problema concreto, la validación implica como paso previo la definición de los requisitos analíticos, o parámetros de calidad (en inglés, *performance characteristics*), que se precisan y, una vez definidos, el siguiente paso consiste en determinarlos.

La primera gran diferencia que encontramos entre el análisis cuantitativo y el análisis cualitativo es la forma de expresar el resultado (figura 2). Es bien conocido que la expresión de un resultado cuantitativo se

caracteriza por dos valores numéricos, el primero es la estimación del valor verdadero y el segundo corresponde a la incertidumbre que está asociada a la dispersión del resultado o intervalo dentro del cual tenemos cierta fiabilidad de que se encuentra el valor verdadero.



Figura 2, Expresión del resultado analítico.

El resultado de un análisis cualitativo también se caracteriza por dos valores, el primero es binario ‘SI/NO’, y el segundo es la probabilidad de error asociada a la decisión tomada.

Parámetros de calidad

Los sistemas de *screening* tienen unas connotaciones especiales que conllevan una cuidadosa adaptación de los parámetros de calidad que están bien definidos y estudiados en los procesos de medida con finalidad cuantitativa, bien sean de tipo físico como químico. En la tabla 2 se muestran los parámetros más relevantes desde el punto de

vista matemático-estadístico, tanto cuantitativos como cualitativos. Algunos de los parámetros de tipo cualitativo, han sido definidos recientemente por la AOAC 1995 [Feldsine 2000], mientras que en otros se está actualmente trabajando en su definición.

<u>Cuantitativo</u>	<u>Binario/Cualitativo</u>
→ Sensibilidad, Especificidad	→ Probabilidad de falso positivo y negativo
→ Selectividad: Interferencias	→ Sensibilidad, Especificidad
→ Límite de detección	→ Selectividad: Interferencias
→ Rango y Linealidad	→ Límite de detección
→ Incertidumbre	→ Límite de corte (cut-off)
→ Exactitud: veracidad, precisión	→ Incertidumbre o región de error
→ Robustez	→ Robustez
→ ...	→

Tabla 2. Parámetros de calidad.

Cabe resaltar que algunos son específicos del análisis cualitativo, como son la proporción de falsos positivos y negativos y el límite de corte o *cut-off*. Otros, como especificidad, sensibilidad y límite de detección, pueden tener un significado ligeramente diferente aunque mantienen el mismo nombre que en el cuantitativo. Finalmente, decir que la mayoría de los parámetros cualitativos, al ser de naturaleza binaria SI/NO, se expresan en términos probabilísticos.

De forma genérica podríamos decir que la validación de cualquier método de análisis implica el establecimiento de los parámetros

de calidad considerados básicos: trazabilidad, exactitud, representatividad, etc. De los parámetros de calidad propios y característicos del análisis cualitativo podemos destacar aquellos que hacen referencia o están relacionados con niveles de concentración; así podemos distinguir el límite de detección, el límite de corte o *cut-off* y el límite legislativo.

La situación habitual que nos solemos encontrar es que la legislación indica el contenido máximo o mínimo de un determinado analito en una muestra. Si tomamos como referencia la presencia de un contaminante, la legislación fija el contenido máximo, por ejemplo el contenido máximo de Aflatoxina B₁ en fruto seco es de 2 ng/g [EC No. 257/2002].

Centrémonos en los métodos cualitativos sensoriales, concretamente en los *test kit* que hoy en día son ampliamente utilizados. Un caso concreto sería aquel en que cuando la muestra contiene menos de 2 ng/g de Aflatoxina B₁ se obtiene una repuesta negativa (aparece color), mientras que si la concentración es superior la respuesta es positiva (no aparece color) [Aflacard B₁ 2002].

En este tipo de *test kits* lo ideal es que el límite real al que criba el *test kit*, denominado límite de corte o *cut-off*, coincida con el límite legislativo, pero así como el legislativo es un límite teórico, el de corte es un límite experimental (figura 3).

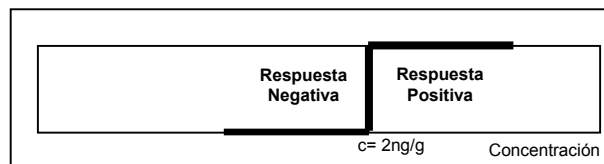


Figura 3, Respuesta teórica de un test kit

Pero desde el punto de vista experimental hay que tener en cuenta que si representamos la respuesta del *kit* en función de la concentración se pueden distinguir 3 zonas (figura 4): a) zona donde se obtiene siempre una respuesta negativa (correctos negativos), b) zona donde para una misma concentración se pueden obtener tanto respuestas positivas como negativas, corresponde a la zona de falsos positivos y negativos, y c) zona donde se obtiene siempre una respuesta positiva (correctos positivos).

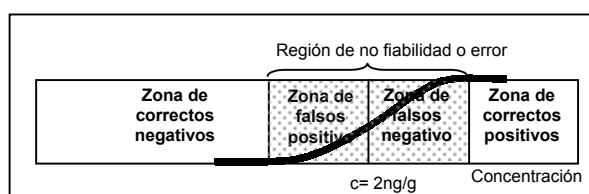


Figura 4, Respuesta experimental de un test kit

En el ejemplo concreto de un contaminante como puede ser la Aflatoxina B₁, está claro que interesa no dar ningún falso negativo (decir que no hay analito cuando en realidad sí lo hay), mientras que sí podemos aceptar falsos positivos puesto que en ningún caso implicaría un riesgo para la salud ya que toda muestra que de una respuesta positiva se analiza posteriormente mediante el método confirmatorio. En este caso concreto, interesa que la zona 2) esté por debajo del límite legislativo para asegurar la ausencia de falsos negativos y además que esté lo más próximo posible al límite legislativo, para reducir el número de falsos positivos, (figura 5).

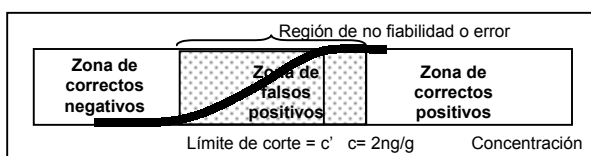


Fig. 5, Límite de corte en el caso de un contaminante

Como queda reflejado en las figuras, alrededor del límite de corte se sitúa una zona o región de error o falta de fiabilidad (que en la literatura inglesa se denomina, *unreliability*). Esta zona corresponde al intervalo de concentraciones donde se obtienen los falsos positivos y negativos, por lo que está definida por un valor superior e inferior de concentración de analito en muestra. Estos dos tipos de errores son dos parámetros básicos en la caracterización de

un sistema de *screening* y se definen como [Massart 1997]:

Falsos negativos. Corresponden a aquellas muestras que contienen uno o más analitos por encima del valor límite permitido (límite legislativo) y que al aplicar el test de *screening* dan una respuesta negativa. Es decir, como resultado del test de *screening* se concluye que la muestra no contiene analito por encima del nivel máximo fijado cuando en realidad sí que contiene.

Falsos positivos. Corresponden a aquellas muestras que realmente no contienen analito por encima del nivel máximo permitido y que sin embargo el test de *screening* indica que están por encima de dicho nivel.

Hay dos parámetros que están relacionados con los límites inferior y superior de concentración que definen la zona de no fiabilidad que son la sensibilidad y la especificidad. Hay que resaltar que, como ya se ha comentado anteriormente, ambos se expresan como probabilidades. La sensibilidad está relacionada con el valor superior de la región de no fiabilidad, mientras que la especificidad lo está con el valor inferior de dicha región.

La **sensibilidad** se define como la capacidad o habilidad del sistema de *screening* de detectar muestras positivas cuando realmente son positivas. De forma similar, la **especificidad** se define capacidad o habilidad del sistema de *screening* de detectar muestras negativas cuando realmente son negativas.

Finalmente, al igual que en los sistemas de medida cuantitativos, debe comprobarse que el método que se utiliza para cribar tenga un límite de detección inferior o igual al límite de corte.

Métodos para caracterizar un sistema de *screening*

Se han desarrollado varios métodos que permiten caracterizar un sistema de *screening*. Cada uno de ellos tiene distinto grado de adaptación a las diferentes situaciones y problemáticas que pueden encontrarse, atendiendo al tipo de medida o resultado obtenido.

Dos de ellos se fundamentan en el concepto de probabilidad y en dicho sentido asocian una probabilidad al resultado de la medida: las tablas de contingencia y el teorema de Bayes [McFall 1999]. Un tercer método se basa en el establecimiento de las curvas

características de funcionamiento [EURACHEM 1999], y finalmente, se describe un método basado en la aplicación de los tests de hipótesis [Pulido 2002]. Este último es similar al que se utiliza para el establecimiento de los parámetros de calidad en el análisis cuantitativo.

A continuación presentamos los rasgos generales de cada una de los métodos sin pretender desarrollarlos, sino dar una idea general de cuando sería más adecuado adoptar cada una de ellos.

Tablas de Contingencia

Las más sencillas son las que diferencian las muestras en dos categorías, positivo/negativo según el método de *screening*, y establece una tabla de comparación respecto al resultado obtenido mediante un método de referencia o confirmatorio (tabla 3).

A partir de la tabla, se calculan los cuatro parámetros básicos, definidos en la sección anterior: falsos positivos, falsos negativos, sensibilidad y especificidad.

		Situación real (método cuantitativo)		
		Igual o superior ●	Inferior ○	Total
Resultado screening	Positivo ●	tp	fp	tp+fp
	Negativo ○	fn	tn	fn+tn
Total		tp+fn	fp+tn	N

Tabla 3, Tabla de contingencia con 2 categorías. Fp: falsos positivos, fn: falsos negativos, tp: total positivos, tn: total negativos, N: número total de ensayos realizados

Una de las ventajas más importantes es su fácil aplicación a múltiples tipos de bioensayos, campo en el que aparecen la mayoría de las aplicaciones [Neil 1975, Hawass 1997].

Cabe resaltar que estos parámetros dan una medida global de la capacidad del método de *screening*. Esto significa que se supone que la muestra problema a examinar se comportará estadísticamente de forma semejante a las ya analizadas, con lo cual no se calcula una probabilidad de error para cada muestra en particular.

Otro de los grandes inconvenientes es que la capacidad de la tabla depende del número de muestras analizadas y el diseño experimental utilizado para elaborar la tabla, y que para el cálculo de los parámetros de calidad mencionados, todas las muestras deben

analizarse por los dos métodos, el de cribado y el confirmatorio.

Teorema de Bayes

El teorema de Bayes se basa en la teoría de probabilidades por lo que, desde el punto de vista analítico, la terminología utilizada es compleja y presenta la dificultad del cálculo de las distintas probabilidades intermedias. En concreto se calcula la probabilidad de dar como válido (positivo o negativo) un resultado cuando en realidad es válido, $P(a/p)$. Esta probabilidad se denomina condicional.

Cabe destacar que una buena estima de la incertidumbre o probabilidad de error mediante este procedimiento implica un número elevado de muestras. Si se compara con las tablas de contingencia, el teorema de Bayes sí permite una estima individual (para cada muestra analizada) de la incertidumbre asociada o probabilidad de dar un resultado erróneo.

Curvas características

Consiste en representar la probabilidad de obtener resultados positivos a distintos niveles de concentración. Esta representación es sigmoïdal, y la posición y amplitud de la

curva es característica de cada sistema de *screening* (figura 6) [Ashley 2001 y 2002].

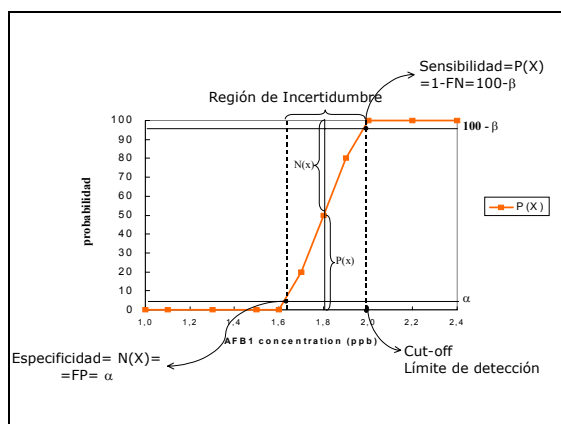


Figura 6, Curva característica de funcionamiento.

El principal inconveniente de este método, al igual que se ha mencionado en el método de Bayes, es la elevada carga experimental ya que hay que realizar un número elevado de análisis a distintos niveles de concentración.

Como ventaja, tal y como se refleja en la figura 6, resaltar que además de los cuatro parámetros básicos (falsos positivos, falsos negativos, sensibilidad y especificidad), se obtiene mucha información adicional sobre las características del método de *screening*: el límite de corte, límite de detección, región de incertidumbre, etc.

Aplicación de los tests de hipótesis

En este caso, se establece el valor de respuesta instrumental al nivel de

concentración al que se quiere cribar (normalmente con un patrón) y para decidir si una muestra problema es positiva o negativa, se compara el valor de respuesta instrumental de la muestra problema con el valor previamente establecido (del patrón). Esta comparación se realiza mediante tests de hipótesis [Pulido 2002].

Como ventajas cabe resaltar la familiaridad que desde un punto de vista analítico se tiene de trabajar con los tests de hipótesis. Si bien la probabilidad α de error (probabilidad de cometer falsos positivos) es mayoritariamente conocida y utilizada en los laboratorios de análisis, no ocurre lo mismo con la probabilidad β (probabilidad de cometer falsos negativos).

Además, es un método rápido y sencillo, fácilmente automatizable, que permite la asignación de una incertidumbre o probabilidad de equivocarnos utilizando como técnica de *screening* una técnica instrumental. Como principal desventaja resaltar que no es directamente aplicable cuando la técnica de *screening* aplicada es un *kit* no instrumental o se basa en una observación visual no cuantificable por parte del analista.

Referencias bibliográficas

P. Feldsine, C. Abeyta and W. Andrews, Journal of AOAC International, 85 (5) (2002) 1187.
URL: <http://aoac.org/vmeth/MicrobiologyGuidelines112700.htm>

Commission Regulation (EC) No 257/2002. Official Journal of the European Community, 12.2.2002, No. L 041, pp12-15.

Aflacard B₁ Product Overview. R-Biopharm Rhone Ltd, 2002.
www.rhone-diagnostics.co.uk

D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbeke. Handbook of Chemometrics and Qualimetrics: Part A. Elsevier, Amsterdam, 1997.

R.M. McFall and T.A. Treat; Annu. Rev. Psychol; 50 (1999) 215-241

EURACHEM. Quantifying uncertainty in analytical measurement, Draft: EURACHEM Workshop, 2nd Edition, Helsinki, 1999.

A. Pulido, I. Ruisánchez, R. Boqué and F.X. Rius. Analítica. Chim. Acta; 455 (2002) 267-275.

B.J. Neil, E. Keeler and S. J. Adelstein; The New England Journal of Medicine; 293 (1975) 211-215.

N. E. Hawass; The British Journal of Radiology, 70 (1997) 360-366.

K. Ashley, R. Song and P. C. Schlecht,. Journal of Hazardous Materials 83 (2001) 29-39.

K. Ashley, R. Song, P. C. Schlecht. American Laboratory 34 (12) (2002) 32-39

Los autores agradecen todos los comentarios relacionados con los contenidos de este artículo. Pueden dirigirse, mediante mensaje electrónico, a la dirección: quimio@quimica.urv.es. Una versión en soporte electrónico de este artículo e información adicional pueden encontrarse en <http://www.quimica.urv.es/quimio>