Artículo Original

Estimación del Valor Predictivo Positivo de la Colangiopancreatografía Magnética utilizando metodos de Bayes.

Luis Bravo-Melo[1], Rafael Tovar-Cuevas[2], Jorge Achcar[3].

[1] Escuela de Estadística, Universidad del Valle Santiago de Cali, Colombia
[2] Estadístico, PhD, Escuela de Estadística, Universidad del Valle, Santiago de Cali, Colombia
[3] Departamento de saude coletiva, Faculdade de Saude de Riberão Preto, Universidade de São Paulo

Fecha de Recepción: 30/06/2015
Fecha de Solicitud de Correcciones: 30/09/2015
Fecha de Aceptación: 15/10/2015

Resumen

Objetivo: este artículo presenta la manera de calcular, dentro de una prueba de diagnóstico clínico, la probabilidad de que un individuo, cuyo resultado para la prueba de tamizaje es positivo, esté realmente enfermo sabiendo que no se tiene su resultado con la prueba patrón de oro; probabilidad denominada Valor Predictivo Positivo (VPP). **Método**: se asumió el VPP como una cantidad desconocida para la que se puede construir una distribución de probabilidad. Dentro del paradigma Bayesiano de la estadística, se utilizó el método propuesto por Winkler and Smith para calcular VPP utilizando un conjunto de registros sobre coledocolitiaisis en pacientes con pancreatitis aguda. **Resultados:** los métodos bayesianos permitieron obtener regiones de credibilidad más estrechas que los intervalos de confianza clásicos. **Discusión:** el uso de métodos Bayesianos es una excelente alternativa para obtener estimaciones más precisas del VPP. Estimar los valores predictivos utilizando los datos de la tabla 2X2 conlleva a errores cuando el tamaño de la muestra no es tan grande como para asumirla cercana a la población.

Palabras Clave: VPP, Sensibilidad, Especificidad, Prevalencia, Elicitación.

Estimation of the Positive Predictive Value of the Magnetic resonance Cholangiopancreatography using Bayes methods

Abstract

Objetive: In this article we present different alternative to compute the Positive Predictive Value (PPV) for a diagnostic test. **Method**: we assumedthe VPP as a continuous and unknown quantity whose natural performance is susceptible of model using a distribution of probability. We use the method propose as Winkler and Smith to estimate the VPP using Bayesian procedure. We illustrate our methodology with a data set of patients with acute pranceatitis taken the Magnetic Cholangiopancreatogrphy as test to screening of Choledocoletise **Results:** The Bayesian approach allows to have a best estimates and the credibility regions were narrower than the confidence intervals.**Discussion:**thje use of Bayesian methods is an excellent choise to obtain more accute estimates. To estimateVPP's using the observed data in table 2X2 implies mistakes when the sample size is not large enough as to think that the sample close the population size.

Keywords. PPV, Sensitivity, Specificity, Prevalence, Elicitation.

Introducción

Tanto en la atención clínica de las personas como en la implementación de programas de salud pública se utilizan criterios o pruebas diagnósticas con el fin de separar a aquellos individuos o poblaciones que tienen el evento de interés en salud de aquellos que no lo tienen. Al decidir qué criterios o pruebas diagnósticas se van a utilizar en el entorno clínico o de salud pública, el médico y el tomador de decisiones deben considerar las características propias de desempeño de los clasificadores como también la capacidad predictiva de los mismos una vez puestos en uso. Tanto a nivel estadístico como en las ciencias de la salud, existe una vasta literatura acerca de los estudios para validar pruebas para diagnóstico, sin embargo, un buen libro que podría ser usado como referencia es el de Sullivan[1]. El Valor Predictivo Positivo (VPP) y el Valor Predictivo Negativo (VPN) como sus nombres lo dicen, son probabilidades "predictivas" es decir, son las proporciones esperadas (no observadas directamente) de individuos realmente enfermos o no enfermos dentro de las respectivas poblaciones de personas con la prueba de tamiz positiva o negativa. En palabras simples, el VPP es la probabilidad de que un individuo cuyo resultado para la prueba tamiz es positivo esté realmente enfermo sabiendo que no tenemos su resultado con la prueba patrón de oro, situación común cuando se aplica la prueba tamiz de manera masiva después de ser validada.

En muchos estudios el tamaño de muestra no es tan "grande" como para considerar que se está cerca de la población de sujetos, lo que conlleva a que la prevalencia del evento de interés no pueda ser directamente estimada usando la tabla de 2X2 en la que se resumen los resultados del estudio. En esas situaciones, se asume la prevalencia como "información externa" a los datos obtenidos en campo que puede estar publicada en alguna otra fuente, razón por la que puede ser interpretada como una información a priori que es actualizada por la información contenida en los datos. En el caso discreto (el trazo o biomarcador está o no presente en el individuo con el evento de interés), el VPP y el VPN toman un valor puntual de probabilidad puesto que se está estimando a partir de valores puntuales de las características de la prueba en evaluación. En el caso continuo, lo que se asume es que el valor predictivo es una variable aleatoria que puede tomar cualquier valor entre los muchos posibles en el intervalo (0,1) y que su distribución en la naturaleza puede ajustarse a una distribución de probabilidad susceptible de ser aproximada de alguna manera. El artículo de Winkler and Smith [2] hace un estudio detallado acerca de esta situación. Existe en la literatura trabajos que presentan estratégias para obtener las estimaciones de los valores predictivos. Smith et al.[3], calculan un VPP estándar con estimaciones aproximadas de la prevalencia, la sensibilidad y especificidad, pero además, analizan el cálculo del VPP con el problema de tener "numerador cero", esto bajo el enfoque beyesiano. Otros trabajos que merecen ser citados por su aporte al cálculo y la importancia del los valores predictivos en estudios de validación de pruebas para diagnóstico clínico, son los de Plumb et al.[4], Meck et al.[5] y Naiditch et al. [6].

En este artículo, se aborda el asunto de estimar los valores predictivos (específicamente el VPP), considerando diferentes escenarios posibles en el procedimiento de estimación que incluyen métodos bayesianos y clásicos de la estadística, y se hace la comparación de los resultados obtenidos. Se utiliza un conjunto de datos de individuos con pancreatitis aguda de origen biliar (PAB) para iliustrar los resultados al asumir la Colangiopancreatografía Magnética (CRM) como una prueba de clasificación inicial y la colangiopancreatografía endoscópica retrograda (CPRE) como patrón de oro.

Datos

En Mogollón et al.[7] se establecen las características de desempeño diagnóstico de la colangiopancreatografía magnética (CRM) temprana en individuos con pancreatitis aguda de origen biliar (PAB) leve para la detección de coledocolitiasis, considerando la colangiopancreatografía endoscópica retrógrada (CPRE) -método invasivo para el diagnóstico de PAB- como el método de referencia para la evaluación. Las características diagnósticas de la CRM en este estudio fueron evaluadas mediante los registros históricos de pacientes de los que se tenía información de haberles realizado CRM y CPRE. De esta manera los autores contaron con 154 historias clínicas de pacientes con PAB leve llevados a CRM, de los cuales 86 contaban con resultado de CPRE y CRM, cantidad que finalmente fue considerada la muestra de estudio.

En un trabajo previo, Tovar[8], obtuvo las estimaciones de la sensibilidad y la especificidad de la CRM utilizando metodología bayesiana de análisis de datos. Para obtener sus estimaciones, asumió que los dos parámetros son variables aleatorias continuas cuyo comportamiento en la naturaleza puede ser modelado usando una distribución de probabilidades teórica (distribución a priori). Se asumió entonces que el comportamiento de cada uno de los parámetros puede ser modelado usando una distribución Beta de probabilidades caracterizada por dos hiperparámetros (cantidades que caracterizan la forma de la distribución). Dado que la distribución a priori expresa la información sobre los parámetros de interés que no está contenida en los datos obtenidos al realizar el estudio de validación en campo (razón por la que se asume que es externa al experimento), entonces se procede a combinar dicha información con la contenida en las observaciones (verosimilitud) para obtener lo que se denomina la distribución a posteriori, la cual no es más que la distribución de probabilidades del parámetro después de combinar la información contenida en el experimento con la información externa al mismo. La distribución a posteriori permite hacer el proceso de estimación a través de alguna de sus medidas de tendencia central y sus percentiles. Existen situaciones en las que no es posible contar con información externa al experimento, como también puede ocurrir que existan diferentes fuentes. Entre las "fuentes externas" más comunmente consideradas están los especialistas en el tema que pueden tener diferentes niveles de experiencia y exposición al evento de interés, resultados de estudios publicados en la literatura, registros médicos de centros de salud, etc. En el caso del estudio de las características de la CRM, se contó con el conocimiento de una medica especialista en cirugía de vías biliares y con artículos publicados previamente en los que se había evaluado el poder clasificador de la CRM para casos de coledocolitiasis. El autor, obtiene las estimaciones bayesianas de la sensibilidad y la especificidad, además del VPP y el VPN para el caso discreto. Ver Tovar[8].

Winkler and Smith[2] desarrollaron un método para calcular el valor predictivo positivo (VPP) para un paciente en particular después de conocer el resultado de la prueba validada como tamiz pero sin haber aplicado el patrón de oro, asumiendo que el VPP es una cantidad aleatoria cuyo valor es incierto y cuya distribución de probabilidad se puede aproximar usando simulación estocástica sencilla.

Metodología de Winkler y Smith distribuciones para VPPs

En su artículo, Winkler y Smith hacen sus estimaciones asumiendo que después de terminado el estudio de tamizaje y validada la prueba se le aplica a un individuo cualquiera de la población de interés si el resutado es positivo, entonces se requiere saber cual es la probabilidad de que el paciente presente realmente el evento de interés. En general, la idea de los autores consiste en pensar en el VPP como una cantidad incierta que refleja la incertidumbre en la prevalencia (p), la sensibilidad (s) y la especificidad (t), por lo que se puede pensar entonces en construir una distribución de probabilidad para los VPPs. Así, mediante la generación de muestras aleatorias de (p,s,t)y calculando

$$q \equiv P(D|+, p,s,t) \qquad (1)$$

donde,

$$P(D|+, p,s,t) = \frac{ps}{ps + (1-p)(1-t)} \qquad (2)$$

para cada muestra se puede construir dicha distribución. Donde D es una variable binaria que representa que el verdadero estado del individuo (enfermo o sano), y el signo (+) simboliza un resultado positivo en la prueba tamiz aplicada para evaluar la presencia de dicha enfermedad. Es posible entonces, calcular la distribución para el VPP usando simulación y a través de la ecuación $P(D|+) = \iiint P(D|+,p,s,t)f(p,s,t|+)$. En la práctica, simular distribuciónes de probabilidad es sencillo si se representa la distribución condicional $f(p,s,t|+)$ como una mezcla (mixtura) de otras dos distribuciones condicionales, las cuales se pueden tener si se sabe que el paciente tiene el evento D o no lo tiene $(\overline{D})$ .Las probabilidades de mezcla son $P(D|+)$ y $P(\overline{D}|+)$:

$$f(p,s,t|+) = f(p,s,t|D,+)P(D|+) + f(p,s,t|\overline{D},+)P(\overline{D}|+). \qquad (3)$$

Para muestrear desde $f(p,s,t|+)$, primero se obtiene $P(D|+)$ usando la ecuación

$$P(D|+) = \frac{E(p)E(s)}{E(p)E(s) + [1-E(p)][1-E(t)]} \qquad (4)$$

ó

$$P(D|+) = \frac{P(D,+)}{P(D,+) + P(\overline{D},+)} \qquad (5)$$

Donde $E(p)$, $E(t)$ y $E(s)$ son las cantidades promedio o esperadas para la prevalencia, la especificidad y la sensibilidad respectivamente (se obtienen despues de simular grandes cantidades de valores y obtener la media de las mismas). Para cada repetición de la simulación se obtiene $P(D|+)$ el cual es un valor entre cero y uno, y luego se simula un número aleatorio uniforme en el intervalo [0,1], de modo que, si el número aleatorio es menor o igual a $P(D|+)$ entonces se asume el resultado positivo del individuo como el de un verdadero caso y extraer (p,s,t) desde $f(p,s,t|D,+)$; de lo contrario, se asume que el individuo con resultado positivo responde a un falso positivo y se extrae (p,s,t) desde $f(p,s,t|\overline{D},+)$. Se obtienen entonces $q \equiv P(D|+,p,s,t)$ de (p,s,t) mediante la ecuación (1) y se repite el proceso un gran número de veces para construir la distribución de q. Finalmente, el valor estimado del VPP será el promedio aritmético de la cadena de valores de q obtenida.

Formas de calcular del VPP a partir de los datos de conteo

Lo más común en los libros de epidemiología y bioestadística que abordan el tema de los estudios de validación de pruebas para diagnóstico, es encontrar métodos de estimación de los valores predictivos usando las cantidades de sujetos que presentaron las combinaciones de resultados de la prueba bajo estudio y la considerada patrón de oro. El estimador se calcula o bien sea usando los datos de la Tabla 2X2 directamente (método que generalmente arroja resultados sesgados) o utilizando la fórmula de Bayes asumiendo conocido el valor de la prevalencia y tomando las estimaciones de la sensibilidad y la especificidad puntuales calculadas a partir de los datos del cuadro. Los procedimientos convencionales para obtener las estimaciones no tienen en cuenta el carácter continuo del indicador y además consideran que el VPP es una probabilidad predictiva de tener el verdadero estado en el caso de que se aplicara la prueba tamiz, es decir, se asume que aun no se sabe el resultado de la prueba tamiz en el sujeto.

Para calcular el VPP se pueden considerar diferentes situaciones, una es que se tienen los resultados de la tabla 2X2 y se calcula directamente de la misma, es decir; sea D la variable aleatoria que identifica el resultado de la prueba patrón de oro en el individuo de modo que D toma el valor de uno cuando el individuo tiene un resultado positivo a esa prueba y toma el valor de cero en el otro caso. Se asume también la variable binaria T que identifica el resultado de la prueba usada para tamiz y que esta bajo validación, los resultados obtenidos después de aplicarle las pruebas a una muestra de sujetos se pueden resumir en una tabla de 2X2 así:

Cuadro 1: Arreglo de datos producto de un estudio de validación de prueba para diagnóstico clínico

	D=1	D=0	Total
T=1	*a*	*b*	*a+b*
T=0	*c*	*d*	*c+d*
Total	*a+c*	*b+d*	*n*

Donde *a* es la cantidad de individuos que presentaron resultado positivo en ambas pruebas, *b* es la cantidad de sujetos con resultado positivo en la prueba bajo validación y resultado negativo en la prueba patrón de oro (falsos positivos), *c* es la cantidad de sujetos que tienen resultado positivo en la prueba patrón de oro y resultado negativo en la prueba tamiz y *d* es el número de personas que comparten un resultado negativo en ambas pruebas.

Una forma de cálculo del estimador del VPP, observada en cursos introductorios de bioestadística o epidemiología es:

$$\widehat{VPP} = \frac{a}{a+b}$$

Esta forma de obtención del estimador presenta problemas cuando el tamaño de la muestra no es lo suficientemente grande como para asumir que tenemos una cantidad de sujetos muy cercana a la población, lo que nos permitiría tener una estimación bastante buena de la prevalencia del evento.

De otro lado, bajo el supuesto de que llega un sujeto que no participó en la etapa experimental, se le aplica la prueba de tamizaje y se observa un resultado (aun no se ha observado, se asume que eso pasaría), se hace necesario estimar la probabilidad condicional en el resultado observado de que el sujeto en cuestión tenga o no realmente el evento. Para el caso de un resultado positivo, se tendría que calcular el VPP que puede escribirse en términos de probabilidades como (ecuación 1):

$$VPP = P(D=1|T=1) = \frac{P(D=1)P(T=1|D=1)}{P(D=1)P(T=1|D=1) + P(D=0)P(T=1|D=0)}$$

Si se sabe que *P(D=1)=prevalencia*; *P(T=1|D=1)=sensibilidad* y *P(T=0|D=0)=Especificidad* , entonces, podemos escribir el VPP como:

$$VPP = \frac{prevalencia*sensibilidad}{prevalencia*sensibilidad + (1-prevalencia)*(1-especificidad)}$$

Revisando los textos de epidemiología y bioestadística, se tiene que la ecuación (1) es la expresión de la fórmula de Bayes, la cual asume que la prevalencia es una probabilidad "externa" que no puede ser calculada a partir de la tabla 2X2 que resume los conteos y debe ser establecida a través de alguna fuente diferente al experimento. Cuando se utiliza este método para obtener el estimador, podemos decir que la prevalencia es una información a priori que se tiene antes de recabar los datos en campo y que debe ser actualizada usando los resultados de los individuos participantes en la investigación (la sensibilidad y la especificidad estimadas).

Cálculo del VPP asumiendo continuidad

En los casos más comunes encontrados en los libros de texto, las fórmulas de cálculo del VPP emplean los valores observados en la tabla de 2X2 y asumen un valor puntual para la prevalencia del evento de interés. Sin embargo, es posible asumir que el VPP es una variable aleatoria que toma valores en el intervalo [0,1] cuyo comportamiento en la naturaleza puede ser modelado usando una distribución de probabilidades. Una distribución adecuada para modelar dicho comportamiento es la distribución de probabilidades Beta con parámetros a y b. La familia de distribuciónes Beta es bastante flexible e incluye un amplio rango de posibles comportamientos al variar sus parámetros dentro de los números reales positivos. Esta forma de asumir la situación es propia de la inferencia Bayesiana, la cual asume que las cantidades que comunmente llamamos parámetros en el ámbito de la inferencia clásica o frecuentista de la estadística, los cuales son asumidos como valores fijos pero desconocidos y propios de la población, en realidad son cantidades aleatorias con un comportamiento probabilístico modelable a través de una distribución teórica.

Dado que el VPP es una función matemática de otras cantidades estimadas como la prevalencia, la sensibilidad y la especificidad, entonces es posible asumir que dichas cantidades también son aleatorias y su comportamiento es susceptible de ser modelado. Claramente, tanto el VPP como la sensibilidad, la especificidad y la prevalencia, son proporciones que en el ámbito poblacional pueden ser vistas como probabilidades así que todos son indicadores que toman valores en el intervalo [0,1].Siendo así, se puede asumir que el comportamiento natural de las cuatro cantidades puede modelarse usando distribuciónes Beta con parámetros a y b diferentes para cada una de acuerdo con las características contextuales en las que se realiza el estudio para validación.

En términos bayesianos, asumimos que tenenos información externa que nos permite obtener valores para a y b en cada caso, de modo que podemos plantear el siguiente modelo para el estudio de validación de una prueba para diagnóstico que dio origen a la ecuación (1) usando la misma notación de Winkler y Smith.

Sea $p = P(D=1)$, sea $s = P(T=1|D=1)$, y sea $t = P(T=0|D=0)$, la prevalencia, sensibilidad y especificidad, de modo que $p \in (0,1)$, $s \in (0,1)$ y $t \in (0,1)$. Si asumimos que el comportamiento de cada cantidad aleatoria puede ser modelado usando una distribución Beta entonces tenemos que:

$$f(p|a,b) = \frac{1}{Beta(a,b)} p^{a-1}(1-p)^{b-1} \qquad a,b > 0\,,\ p \in (0,1)$$

Donde $Beta(a,b)$ es una constante llamada de normalización que permite que la integral de la función sea igual a uno. La función de probabilidad para modelar el comportamiento de la sensibilidad y la especificidad tendrán la misma forma con la diferencia de que a y b son únicos para cada una de las cantidades aleatorias. En el artículo de Tovar[8] se presenta en forma detallada el procedimiento para obtener las estimaciones de la sensibilidad y la especificidad, mediante métodos bayesianos y asumiendo una distribución Beta para modelar la información a priori o externa.

En el proceso de estimación usando metodología bayesiana, puede ocurrir que se tengan diferentes fuentes externas para obtener información sobre la cantidad aleatoria de interés. Es posible contar con personas consideradas especialistas en el tema que pueden brindar información desde su conocimiento y exposición al evento de interés, o puede ser que se cuente con documentos publicados como artículos científicos o reportes técnicos, entre otras posibles opciones. Cuando se tiene este tipo de situaciones, se tendrán tantos modelos como posibles fuentes de información, caso en el que se hace necesario escoger entre las diferentes alternativas la que mejor se ajuste a la situación real. Existen diferentes estrategias teóricas creadas para seleccionar el modelo que mejor explique el fenómeno real, entre ellas el factor de Bayes. En el artículo de Tovar[8] se puede encontrar una explicación más detallada sobre los métodos de selección de modelos.

Estimación del valor predictivo de la Colangiopancreatografía magnética (CRM)

Usando la Colangiopancreatografía endoscópica retrógrada (CPRE) como prueba patrón de oro, se estimó el VPP de la CRM utilizando los datos de conteo y la fórmula de Bayes para el caso discreto y utilizando la metodología desarrollada por Winkler y Smith para el caso continuo.

Situación 1: Estimación a partir de la tabla de 2X2

Para llevar a cabo el procedimiento de estimación se contaba con una muestra de 86 individuos diagnosticados con pancreatitis aguda leve y tratados en un hospital público de Colombia. De acuerdo con los resultados reportados por Mogollón et al.[7], en ese estudio, los valores de los conteos en la tabla cruzada fueron: $a = 68, b = 9, c = 2$ y $d = 7$ de modo que la estimación del VPP usando unicamente los datos de la Tabla 1 sería $\widehat{VPP} = \frac{68}{77} = 0.88$ lo que implicaría que la prevalencia estimada para la coledocolitiasis sería de $\hat{p} = \frac{70}{86} = 0.81$

Situación 2: estimación usando la fórmula de Bayes

En este caso, se consultó a la especialista quien reportó que de acuerdo con los estudios y los datos reportados por el ministerio de salud de Colombia, la prevalencia de coledocolitiasis entre individuos con pancreatitis aguda leve es del 20%. Usando este dato y los valores de prevalencia estimados con la Tabla 2X2 se tiene que:

$$Sensibilidad\ de\ la\ CRM = \frac{68}{70} = 0.97$$

$$Especificidad\ de\ la\ CRM = \frac{7}{16} = 0.44$$

$$\widehat{VPP}\ de\ la\ CRM = \frac{0.20*0.97}{0.20*0.97 + 0.80*0.56} = 0.30$$

Situación 3: estimación del VPP mediante la formula de Bayes usando el valor puntual de la prevalencia y estimaciones bajo continuidad de la sensibilidad y la especificidad.

Tovar[8] en su artículo utilizó tres modelos de estimación para cada uno de los parámetros de desempeño de la CRM. Primero estimó asumiendo a priori que todos los valores de las cantidades en el intervalo (0,1) tenían la misma probabilidad de ocurrir, es decir, utilizó una distribución de probabilidad Uniforme (0,1) como expresión de la información previa a los datos del estudio, de modo que los resultados reportados y el VPP estimado son: $\hat{s} = 0.958$ y $\hat{t} = 0.44$ y $\widehat{VPP}\ de\ la\ CRM = 0.299$. En un segundo caso, expresó la información a priori contenida en artículos científicos publicados previamente sobre el asunto a través de una distribución Beta con parámetros $a = 767.6$ y $b = 80.6$ para la sensibilidad y $a = 83.7, b = 16.5$ para la especificidad. Asumiendo la prevalencia del 20%, los resultados observados fueron: $\hat{s} = 0.91$ $\hat{t} = 0.78$, y $\widehat{VPP}\ de\ la\ CRM = 0.508$. En el caso tres, las estimaciones se hicieron tomando información subjetiva entregada por la especialista y usando distribuciónes Beta para modelar la misma. En este caso, los valores de los hiperparámetros y de las estimaciones fueron: $a = 866, b = 152{,}9$ para la sensibilidad, $a = 48{,}9, b = 6.99$ para la especificidad, $\hat{s} = 0.86$, $\hat{t} = 0.78$ y $\widehat{VPP}\ de\ la\ CRM = 0.49$

Situación 4: estimación del VPP mediante la fórmula de Bayes asumiendo continuidad en los parámetros de desempeño y la prevalencia

Dado que Tovar[8] no estimó la prevalencia de forma continua y para sus estimaciones de la sensibilidad y la especificidad utilizó el valor entregado por la especialista (20%), en esta parte del trabajo, se asumió que, de acuerdo con la naturaleza de este indicador, su comportamiento puede ser modelado usando una distribución Beta de probabilidades y para obtener los parámetros de la distribución a priori (hiperparámetros) se aplicó el procedimiento desarrollado por Tovar[9], definiendo a priori que hay probabilidad de 0.95 de que la prevalencia tome valores entre 0.10 y 0.30 dentro del intervalo (0,1). Siendo así, se tendría una probabilidad de 0,05 de encontrar valores de prevalencia fuera de ese rango. Dado que solo se tiene un intervalo en el que se puede encontrar el valor real del indicador, se utilizó el teorema de Chebychev para obtener una varianza aproximada del comportamiento de la cantidad aleatoria (la prevalencia) dentro del intervalo establecido y empleando un sistema simple de dos ecuaciones con dos incognitas se obtuvo el valor de los hiperparámetros. La distribución posterior de la prevalencia será entonces una $Beta(80.8{,}324.2)$ así que la estimación de la prevalencia será $p = 0.205$. Tomando ese valor junto con las estimaciones de los parámetros de desempeño utilizados en la situación tres, se obtuvierón las siguientes estimaciones para el VPP de la CRM: $\widehat{VPP} = 0.3061$; $\widehat{VPP} = 0.5068$; $\widehat{VPP} = 0{,}49$

Situación 5: estimación del VPP usando método de Winkler y Smith asumiendo una distribución a priori de referencia para el VPP, las estimaciones bayesianas de la sensibilidad y la especificidad y la prevalencia obtenida con los conteos de la tabla 2X2.

Dado que los datos en las celdas de la Tabla 1, corresponden a cantidades de sujetos con resultados de las pruebas de clasificación, se puede aproximar el modelo de probabilidad general para todas las observaciones de la tabla con una distribución multinomial. Sin embargo, cuando se van a calcular las estimaciones de la sensibilidad y la especificidad, se puede partir del supuesto de que se tiene primero el resultado de la prueba patrón de oro y luego se sabe el resultado de la prueba para validación. Partiendo de esta suposición se puede asumir que la distribución de los resultados de la prueba se ajusta

a una distribución binomial de probabilidades con parámetros $s = sensibilidad$ y $n = a + c$ siendo a y c las cantidades de sujetos observadas en la Tabla 1. De igual manera para la especificidad, se tendría que la distribución del número de individuos con resultado negativo en la prueba tamiz, se distribuye binomial con parámetros $t = especificidad$ y $n = b + d$. De igual forma, la cantidad de sujetos con patrón de oro positivo entre todos los individuos examinados, se puede ajusar a una distribución binomial de probabilidades con parámetros $p = prevalencia$ y $n = a + b + c + d$. Volviendo a la estimación de los parámetros de la CRM, se obtuvo la distribución posterior para la prevalencia asumiendo que todos los valores posibles de la cantidad aleatoria tienen la misma probabilidad de ocurrir en el intervalo (0,1), es decir, se asume que la distribución a priori de la prevalencia es una Uniforme (0,1) que es lo mismo que una Beta(a=1, b=1). Entonces se puede escribir la función de verosimilitud de los datos para la prevalencia como:

$$L(p|70) = \binom{86}{70} p^{70}(1\text{-}p)^{16}.$$

Dado que se asumió una distribución a priori uniforme, entonces se tiene que la distribución posterior será una $Beta(71,17)$. Usando las estimaciones de la sensibilidad y la especificidad, se aplica el procedimiento de Winkler y Smith y se tiene que la distribución de probabilidad conjunta $f(p,s,t)$ es un producto de tres densidades Beta:

$$f(p,s,t) = f_\beta(p|71,17) f_\beta(s|69,3) f_\beta(t|8,10),$$

De modo que:

$$E(p) = \frac{71}{88}; \quad E(s) = \frac{69}{72}; \quad E(t) = \frac{8}{18}.$$

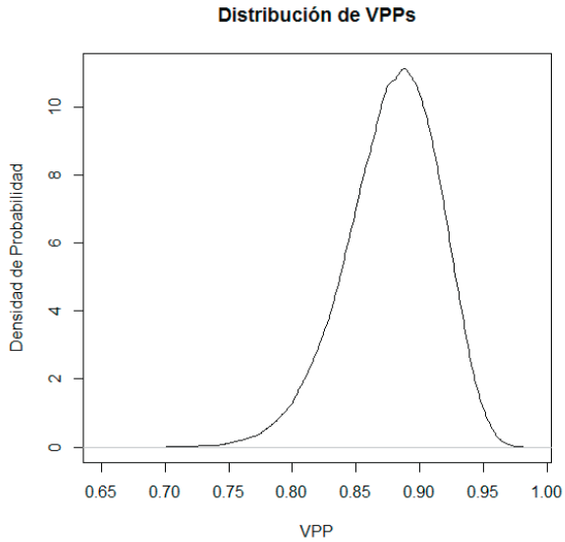
Con lo cual, haciendo uso de la ecuación (4) en el proceso de Winkler and Smith se puede calcular la probabilidad para un individuo, esto es $P(D|+) = 0.878$. Así, observando una lectura positiva en la prueba, pero sin saber si el individuo padece coledocolitiasis, la distribución de probabilidad conjunta $f(p,s,t|+)$, es una mezcla de dos densidades de la forma:

$$\begin{aligned} f(p,s,t|+) = {} & 0.878 f_\beta(p|72,17) f_\beta(s|70,3) f_\beta(t|8,10) \\ & + 0.122 f_\beta(p|71,18) f_\beta(s|69,3) f_\beta(t|8,11) \end{aligned}$$

En donde el primer término de la expresión considera la posibilidad que el individuo tenga coledocolitiasis, por lo que incrementa la cuenta para la prevalencia (de 71 casos de coledocolitiasis pasa a 72), y de igual forma para la sensibilidad (de 69 casos correctamente identificados pasa a 70). Asimismo, el segundo término corresponde a considerar que el individuo no tiene coledocolitiasis, con lo cual se aumenta la cuenta de los pacientes que no padecen esta enfermedad (de 17 pasa a 18), y aumenta el número de pacientes sin coledocolitiasis que han sido categorizados incorrectamente (de 10 pasa a 11).

Continuando con el proceso, se simulan100.000 valores de $f(p,s,t|+)$ y se obtiene el valor de $q \equiv P(D|+,p,s,t)$ para cada uno de los valores simulados, lo que permite obtener la forma de la densidad empírica (con datos y no una ecuación) del VPP. Ver Figura1.

Figura 1: Distribución de los valores del VPP para CRM asumiendo distribuciónes a priori uniformes para la prevalencia, la sensibilidad y la especificidad. Elaboración propia en software R.



Con miras a resumir la información contenida en la distribución hallada, se puede obtener valores de ciertos indicadores como la media de las estimaciones de los VPPs, su error estándar (desviación estándar de las estimaciones) y un intervalo de credibilidad dentro del cual se espera que se encuentre el 95% de los valores del VPP si se le aplicara la CRM a una población de individuos que no fueron evaluados con la CPRE. Los resultados obtenidos son:

$VPP = E(\widehat{VPP}) = 0.8779$; $EE(\widehat{VPP}) = 0.0366$; $ICr_{95\%}(VPP) = [0.797; 0.940]$.

Situación 6: estimación del VPP de la CRM asumiendo que se tiene una prevalencia del 20% y las estimaciones Bayesianas de los parámetros de desempeño obtenidas con distribuciónes uniformes para expresar la información a priori

Para este caso, asumimos que la prevalencia de coledocolitiasis en la población de individuos con pancreatitis aguda leve es del 20% como lo reportó la especialista, de modo que si se tienen 86 individuos en la muestra se esperaría que 17 de ellos tuvieran coledocolitiasis. Tomando este dato se aplica de nuevo el procedimiento realizado en la situación anterior y se tiene que la distribución posterior de la prevalencia después de asumir un comportamiento previo uniforme, es una $Beta(18,70)$. Después de usar los valores de los hiperparámetros usados por Tovar[8], se tiene que $f(p,s,t) = f_\beta(p|18,70) f_\beta(s|69,3) f_\beta(t|8,10)$ de modo que:

$$E(p) = \frac{18}{88}; \quad E(s) = \frac{69}{72}; \quad E(t) = \frac{8}{18}.$$

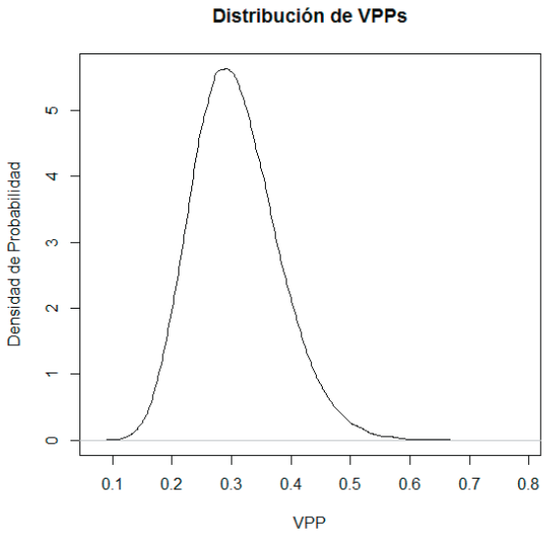
Utilizando los valores obtenidos es posible obtener la probabilidad de que un individuo tenga coledocolitiasis dado que se observó un resultado positivo en la CRM, esto es $P(D|+) = 0.307$ Así, contando con un resultado positivo de la CRM sin saber si el individuo padece coledocolitiasis, la distribución de probabilidad conjunta $f(p,s,t|+)$, es una mezcla de dos densidades:

$$\begin{aligned} f(p,s,t|+) = {} & 0.307 f_\beta(p|19,70) f_\beta(s|70,3) f_\beta(t|8,10) \\ & + 0.693 f_\beta(p|18,71) f_\beta(s|69,3) f_\beta(t|8,11) \end{aligned}$$

Después de realizar 100.000 simulaciones de la distribución encontrada y se obtiene igual cantidad de valores de q lo que permite

establecer la distribución del VPP de la CRM bajo las condiciones dadas. Figura 2

Figura 2: Distribución de los valores del VPP para CRM asumiendo las condiciones establecidas para la condición cinco Elaboración propia en software R.



Las estimaciones obtenidas en este caso son: $VPP = E(\widehat{VPP}) = 0.3074$; $EE(\widehat{VPP}) = 0.0720$; y $ICr_{95\%}(VPP) = [0.183; 0.464]$

Situación 7: estimación del VPP de la CRM asumiendo las estimaciones bayesianas basadas en distribuciónes a priori construidas a partir de información publicada en artículos científicos

Se aplicó el procedimiento de estimación de nuevo, pero esta vez, tomando los valores de la sensibilidad y la especificidad obtenidos a través de la combinación de los conteos con la información contenida en estudios para validación de la CRM realizados en otros países y la estimación de la prevalencia considerada en la situación cuatro. Con base en el intervalo establecido a priori para la prevalencia (0.1, 0.3) se obtuvo después de usar el procedimiento de Tovar[9] una distribución a priori $Beta(63.8,255.2)$ de modo que al combinarla con la información contenida en los conteos de la tabla de 2X2 se obtuvo una distribución posterior $Beta(80.8,324.2)$. Por tanto;

$$f(p,s,t) = f_\beta(p|80.8,324.2)f_\beta(s|822.6,96.6)f_\beta(t|86.7,22.5).$$

Lo que conlleva a que:

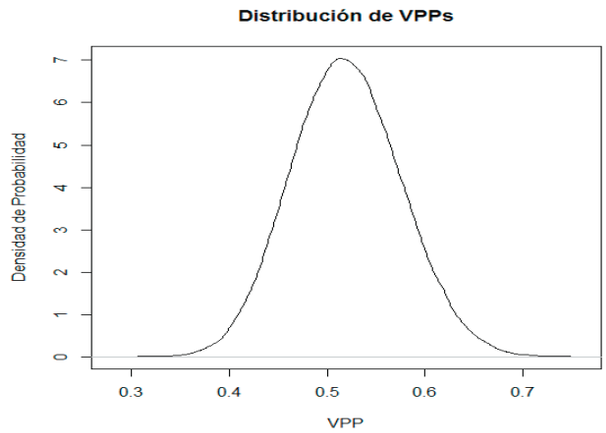
$$E(p) = \frac{80.8}{405}; \quad E(s) = \frac{822.6}{919.2}; \quad E(t) = \frac{86.7}{109.2}$$

Se obtuvo la probabilidad $P(D|+) = 0.5198$. para obtener luego, la distribución de probabilidad conjunta $f(p,s,t|+)$, como una mezcla de dos densidades así:

$$f(p,s,t|+) = 0.52f_\beta(p|81.8,324.2)f_\beta(s|823.6,96.6)f_\beta(t|86.7,22.5) + 0.48f_\beta(p|80.8,325.2)f_\beta(s|822.6,96.6)f_\beta(t|86.7,23.5)$$

Después de hacer 100.000 simulaciones, se obtiene la distribución para los VPPs. Figura 3.

Figura 3: Distribución de los valores del VPP para CRM asumiendo las condiciones establecidas para la situación siete. Elaboración propia en software R.



En este caso, las estimaciones obtenidas son: $VPP = E(\widehat{VPP}) = 0.52$; $EE(\widehat{VPP}) = 0.06$; y $ICr_{95\%}(VPP) = [0.41; 0.63]$

Situación 8: estimación del VPP de la CRM asumiendo las estimaciones bayesianas basadas en distribuciónes a priori construidas a partir de información suministrada por un especialista en el tema

El procedimiento se repitió de nuevo, pero esta vez cambiando las distribuciónes Beta a priori que expresaban los conocimientos previos acerca de la distribución natural de la sensibilidad y la especificidad, asumidas ambas como cantidades con un comportamiento aleatorio en la naturaleza (Ver Tovar[8] para detalles). La prevalencia fue estimada de forma continua asumiendo que la misma se mueve en un rango de (0.1, 0.3) con una probabilidad de 0.95 y que la probabilidad de encontrar valores fuera de ese rango es de 0.05. Se tiene entonces que:

$$f(p,s,t) = f_\beta(p|80.8,324.2)f_\beta(s|921.0,168.9)f_\beta(t|51.9,12.9).$$

Así que:

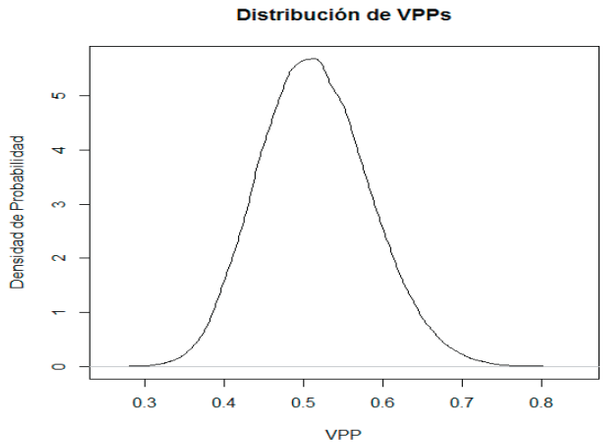
$$E(p) = \frac{80.8}{405}; \quad E(s) = \frac{921}{1089.9}; \quad E(t) = \frac{51.9}{64.8}$$

Con lo que, la probabilidad para un individuo es $P(D|+) = 0.514$ y la distribución de probabilidad conjunta para las tres cantidades de interés después de haber observado un resutado positivo en la CRM es:

$$f(p,s,t|+) = 0.514f_\beta(p|81.8,324.2)f_\beta(s|922,168.9)f_\beta(t|51.9,12.9) + 0.486f_\beta(p|80.8,325.2)f_\beta(s|921,168.9)f_\beta(t|51.9,13.9)$$

Después de simular 100.000 valores de $f(p,s,t|+)$, y encontrar los respectivos valores de q se tiene que la forma de la densidad para los VPPs es la que aparece en la Figura 4.

Figura 4: Distribución de los valores del VPP de la CRM asumiendo las condiciones establecidas para la situación ocho. Elaboración propia en software R.



Finamente, para este modelo, se tienen las siguientes estimaciones: $VPP = E(\widehat{VPP}) = 0.514$; $EE(\widehat{VPP}) = 0.068$; y $ICr_{95\%}(VPP) = [0.387; 0.656]$

Discusión

El Valor predictivo de una prueba para clasificación clínica, es una característica muy importante que permite estimar la probabilidad condicional del verdadero estado de un individuo después de conocer el resultado clasificador de la prueba. De acuerdo con Sullivan[1] los valores predictivos (positivo y negativo) cuantifican el valor clínico de la prueba de tamizaje. De otro lado, según Silva[10], la sensibilidad y la especificidad son medidas de la eficiencia del citerio clasificador que expresan la calidad del comportamiento del mismo ante la presencia o ausencia del evento de interés pero no necesariamente de su utilidad para pronunciarse en enclaves prácticos reales sobre si un sujeto está o no enfermo. El individuo susceptible de ser clasificado (generalmente un paciente) y el profesional de la salud, pueden estar más interesados en la probabilidad de tener el evento de interés dado que el resultado de la prueban tamiz fue positivo. Un aspecto importante es que el valor predictivo de la prueba depende de la cantidad de individuos que presentan el evento de interés en la población de estudio, mientras que la sensibilidad y la especificidad son características inherentes a la estructura de la prueba y su capacidad clasificadora. Es posible entonces, tener una prueba para tamizaje con una alta sensibilidad, pero si la cantidad de individuos enfermos o infectados en la población es muy baja (prevalencia tendiente a cero) el valor predictivo de esa prueba será muy bajo también pues muchos de los individuos que sean positivos a la misma serán falsos positivos dada la dificultad de encontrar verdaderos casos en la población.

Generalmente, en los estudios de validación de pruebas para clasificación diagnóstica, se calcula el valor predictivo a partir de los datos discretos obtenidos luego de los conteos conseguidos después de aplicar el nuevo criterio clasificador entre individuos que se saben tienen la característica de interés, o entre individuos a los que se les aplica un procedimiento clasificador que se asume libre de error. Para hacer la estimación a través de los datos discretos, se emplean dos estrategias:

1. Calcular el VPP a partir de los datos del areglo 2X2 obtenido después de hacer los conteos. Esta estrategia presenta bastantes problemas desde el punto de vista estadístico que conllevan a fuertes problemas desde el punto de vista de la salud pública. Si se asume que la cantidad de personas incluidas en la evaluación de la prueba en campo es lo suficientemente grande como para acercarse a la población, entonces se puede estimar la prevalencia del evento de interés a través de los conteos en la tabla 2X2, siempre y cuando los datos en campo se hayan colectado bajo un diseño de cohorte. Sin embargo, el contar con miles de individuos solo es posible cuando se realizan tamizajes poblacionales y antes de llegar a estos, se supone que la prueba ha sido validada con un grupo menor de sujetos. En la realidad, es difícil asumir que se pueden realizar estudios de evaluación de pruebas en los que participan cientos o miles de personas debido a las condiciones logísticas para desarrollarlos y específicamente a los costos que tales estudios implican. El diseño de estudio de cohorte, exige contar con una prueba considerada patrón de oro que sea aplicada simultáneamente a los sujetos junto con la prueba en evaluación, lo cual garantiza aleatoriedad en los cuatro totales de la tabla 2X2 de modo que el único total fijo, o conocido de antemano, es la cantidad total de sujetos participantes. Cuando los datos son colectados bajo este diseño, es posible estimar la prevalencia del evento de interés a partir del cociente entre el total marginal de la columna con patrón de oro positivo y la cantidad total de participantes en el estudio. Es importante aclarar que lo deseable es tener una muestra de individuos participantes lo suficientemente grande como para obtener una estimación de la prevalencia a la que le apliquen las características propias de un buen estimador.
2. Calcular el VPP usando la formula de Bayes. Esta estratégia metodológica implica que la prevalencia del evento de interés puede ser obtenida en alguna fuente secundaria externa a los datos de los conteos. Algunos diseños para obtener los datos en campo, como el de tener un grupo de individuos con la característica de interés (grupo de enfermos por ejemplo) y un grupo de individuos sin la misma o con otras características que comparten rasgos de similaridad con la de interés pero que no son exactamente ésta. En esos estudios los totales de sujetos son fijos o conocidos por el investigador antes de obtener los datos de conteo, así que, para calcular el VPP necesariamente se debe consultar otras fuentes que permitan tener un valor para la prevalencia.

Generalmente, los valores predictivos son calculados a partir de los datos discretos y con base en la fórmula de Bayes que actualiza la probabilidad a priori (prevalencia) con la información contenida en los datos colectados en campo. Aun cuando este es un procedimiento válido y legítimo, es claro que en su aplicación se puede estar perdiendo información importante acerca de los indicadores puesto que se está trabajando con valores puntuales. Asumir continuidad, fortalece el proceso de estimación, pues, al hacer esto, se está considerando todo el rango de posibles valores que puede llegar a tomar el indicador el cual es resumido a través de su media aritmética. De acuerdo con los resultados obtenidos en el presente trabajo, es posible concluir que calcular el VPP a partir de los datos de la tabla, puede sobre estimar el indicador independiente de que el procedimiento de estimación se haga solo con datos discretos o considerando la continuidad. Es importante tener en cuenta que, los desarrollos y procesos expuestos en este trabajo no son exclusivos para obtener el valor predictivo positivo (VPP) ya que, los mismos, también aplican de manera análoga, para obtener los valores predictivos negativos (VPN).

Agradecimientos:

Los autores agradecen al editor y los evaluadores por sus valiosos comentarios que permitieron mejorar el texto.
Este artículo es parte del trabajo de Maestría en Estadística del primer autor. La participación del segundo y tercer autor fue parcialmente financiada por la Coordenação de aperfeiçoamento de pessoal de nivel superior (CAPES) de Brasil.

Conflicto de intereses:

Los autores declaran que no tienen conflicto de intereses.

Referencias

1. Sullivan M. The statistical evaluation of medical tests for classification and prediction. Oxford University Press. 2003; pp. 22-25.

2. Winkler R, Smith J. On uncertainty in medical testing. Medical Decision Making. 2004; 24: 654-658.

3. Smith J, Winkler R,Fryback D. The First Positive: Computing Positive Predictive Value at the Extremes. American Society of Internal Medicine. 2000; 132:804-809.

4. Plumb A, Taylor S, Halligan S. Effect of faecal occult blood positivity on detection rates and positive predictive value of CT colonography when screening for colorectal neoplasia. Clinical Radiology. 2015May 28; ISSN: 0009-9260.

5. Meck, J.M. et al. Research: Noninvasive prenatal screening for aneuploidy: positive predictive values based on cytogenetic findings. American Journal of Obstetrics and Gynecology. 2015Apr. 1; ISSN: 0002-9378.

6. Naiditch, J.A, Barsness, K.A. The positive and negative predictive value of transabdominal color Doppler ultrasound for diagnosing ovarian torsion in pediatric patients.Journal of Pediatric Surgery. 2013 June 1; 48, 1283-1287, ISSN: 0022-3468.

7. Mogollón G, Sefair C, Upegui D, Tovar J. Colangiopancreatografía magnética: valor diagnóstivo para detectar coledocolitiasis en pacientes con pancreatitis aguda leve. Revista Cubana de Cirugía. 2014; 53(1), pp. 41-51.

8. Tovar J. Inferencia Bayesiana e Investigación en salud: un caso de aplicación en diagnóstico clínico. Rev. Méd. Risaralda. 2015; 21(1), pp.9-16.

9. Tovar J. Eliciting Beta prior distributions for binomial sampling. Rev. Bras. Biom. 2012; v.30, n.1, p.159-172.

10. Silva L. Cultura estadistica e investigacion cientifica en el campo de la salud: una mirada crítica. Editorial Diaz de Santos S.A. 1997; pp. 250-257.