

Inferencia Bayesiana e Investigación en salud: un caso de aplicación en diagnóstico clínico

José Rafael Tovar Cuevas,

Estadístico, PhD. Escuela de Estadística, Facultad de Ingenierías Universidad del Valle, Cali Colombia.
Correo electrónico jose.r.tovar@correounivalle.edu.co

Fecha de Recepción: 5/02/2014

Fecha de Solicitud de Correcciones: 10/19/2014

Fecha de Aceptación: 02/05/2015

Resumen

Introducción: Los métodos propios de la inferencia estadística bayesiana son una alternativa bastante útil y confiable para realizar los análisis de datos recogidos dentro de los proyectos de investigación en las diferentes áreas de la salud. Una de las aplicaciones más comunes se da en los estudios de validación de pruebas para diagnóstico clínico, en los que el uso del teorema de Bayes para la obtención de los valores predictivos de la prueba en evaluación es el procedimiento natural. **Materiales y métodos:** Se desarrolló un procedimiento para estimar la sensibilidad y la especificidad de la colangiografía magnética como prueba tamiz para clasificar individuos con coledocolitiasis. El análisis de datos incluyó obtener los indicadores de desempeño de la colangiografía utilizando metodología bayesiana tanto en el caso discreto como en el caso continuo. **Resultados:** Tanto para la sensibilidad como para especificidad, las estimaciones bayesianas presentaron un mayor grado de precisión frente a las estimaciones obtenidas mediante el método clásico de máxima verosimilitud. **Conclusión:** los métodos bayesianos son bastante útiles en situaciones como la tratada en este artículo en la que los tamaños de muestra para obtener las estimaciones pueden considerarse muy “pequeños” lo cual hace que las estimaciones usando métodos no bayesianos no sean tan confiables debido a su poca precisión.

Palabras clave: inferencia bayesiana, diagnóstico clínico, sensibilidad, especificidad, teorema de Bayes

Bayesian inference and health research studies: an application case in clinical diagnosis

Abstract

Introduction: the bayesian methods are a helpful and reliable alternative to analyze the data obtained in research Projects within the health sciences. One of the most common applications is given in the clinical diagnostic tests validation studies, in which the use of Bayes' theorem to obtain predictive values of the test under evaluation is the natural choice. **Material y methods:** A method to estimate the sensitivity and specificity of magnetic resonance cholangiography as a screening test to classify individuals with choledocholithiasis was developed. The estimation process was developed using the Bayesian methodology. **Results:** The Bayesian estimates of the sensitivity and specificity had a greater degree of accuracy compared to the estimates obtained by the classical method of maximum likelihood. **Conclusion:** Bayesian methods are quite useful in situations such as discussed in this article that the sample sizes for the estimates can be considered very “small”. In those situations the results obtained using non Bayesian methods may be more inaccurate and unreliable. **Key words:** Bayesian inference, Clinical diagnosis, sensibility, specificity, Bayes theorem.

Introducción

Muchos de los desarrollos de la estadística y básicamente de los métodos estadísticos, han estado fuertemente ligados a los problemas que se presentan en las diferentes áreas del conocimiento relacionadas con la salud de los seres humanos. Buena parte de los métodos estadísticos conocidos tuvieron su origen en el planteamiento de propuestas de análisis a los datos obtenidos en el desarrollo de medicamentos, esquemas de tratamiento para enfermedades o estudios encaminados al conocimiento de los procesos presentes cuando los individuos están expuestos a procesos de enfermedad o infección. Durante muchos años, los procedimientos desarrollados para el análisis de los datos obtenidos bajo los esquemas y protocolos desarrollados por las ciencias médicas, se han enmarcado dentro del contexto del paradigma clásico o frecuentista de la estadística, el cual, toma las mediciones realizadas en un conjunto de individuos (muestra) que cumplen con las condiciones requeridas para estudiar el asunto o fenómeno de interés y les aplica procedimientos aritméticos y de cálculo bajo el soporte teórico de las matemáticas, con el fin de obtener información relevante que permita validar el modelo seleccionado por los investigadores como un buen reflejo del mecanismo generador de los cambios en el estado natural libre de enfermedad o infección. De acuerdo con Silva y Benavides, si bien la estadística 60 años atrás, prácticamente no se utilizaba en las investigaciones médicas, en este momento muchos investigadores del ámbito salubrista, epidemiológico o clínico sienten que sus trabajo no tienen suficiente rigor científico si no vienen avalados por al menos un procedimiento estadístico. Según los mismos autores, la tarea más importante de la estadística es proporcionar alternativas cuantitativas objetivas al mero juicio personal, que permitan una interpretación adecuada de los datos producidos por los estudios (1). Aun cuando el uso de los métodos estadísticos clásicos es cada vez más común en la investigación médica, es interesante ver como el mismo se ha convertido también con el tiempo, en un proceso mecánico y muchas veces falto de rigor científico utilizado por los investigadores con el fin de conseguir colocar sus resultados en un lenguaje aceptado por la comunidad científica internacional pero desconociendo el transfondo teórico matemático y filosófico que soporta la creación del método y su utilidad.

En las últimas décadas y con el avance en los desarrollos tecnológicos computacionales, se ha tenido un fuerte avance en el desarrollo de los métodos propios de la inferencia bayesiana, corriente de la estadística que no entiende la probabilidad como la frecuencia relativa de un suceso a largo plazo, sino como el grado de convicción personal acerca de que el suceso ocurra o puede ocurrir, lo que se conoce como definición subjetiva de probabilidad (2). La inferencia bayesiana se soporta en la idea simple de que la única descripción satisfactoria de la incertidumbre es por medio de la probabilidad de modo que las inferencias (extracción de información y conocimiento acerca del fenómeno en estudio) se realizan utilizando probabilidades que combinan el conocimiento subjetivo ganado por un individuo que ha estado expuesto previamente al fenómeno y que no se encuentra cuantificado en forma de datos, con la información proporcionada por un conjunto de mediciones obtenidas bajo condiciones experimentales. Algunos autores como Gill et al (3) afirman que los médicos son bayesianos naturales que utilizan el razonamiento bayesiano de manera natural específicamente cuando deben diagnosticar a un individuo de acuerdo con su cuadro de síntomas (Aun cuando este documento

se puede centrar en los médicos como profesionales de la salud, es importante especificar que no solo ellos realizan el acto de clasificar individuos dentro de un grupo con o sin una característica de salud a partir de la información brindada por el individuo consultante). Cuando una persona consulta por problemas relacionados con su salud informa al profesional sobre los aspectos que considera no están dentro de la normalidad con su cuerpo y a su vez la persona que escucha hace preguntas para refinar la información que está recibiendo. Durante este proceso, el profesional de la salud procesa mentalmente la información y aunque no cuantifica directamente, asigna una probabilidad o posibilidad a que el cuadro que está siendo presentado por el consultante puede estar más relacionado a una condición de salud que a otras, lo que le permite tomar la decisión de pedirle a la persona consultante resultados de exámenes o pruebas cuyo resultado “objetivo” confirmará o desvirtuará su apreciación inicial, es decir, el resultado de las pruebas actualizará la información subjetiva procesada a partir de los síntomas y aumentará o disminuirá la probabilidad inicial de que la persona consultante presente la condición de salud intuida por el profesional de la salud. Saliendo del contexto de atención personalizada y entrando en procesos de investigación y desarrollo de conocimiento respecto a problemas relacionados con la salud de los individuos, el razonamiento bayesiano ha tomado gran fuerza como una alternativa para la obtención de los resultados de los proyectos. Algunos autores como Silva (4), Ashby y Smith (5) y Ashby (6) se han concentrado en evaluar la utilidad del abordaje bayesiano en estudios del área de la salud, muchos han trabajado en la aplicación de los mismos en la evaluación de métodos para diagnóstico clínico, sin embargo, considerando el alcance de este artículo, es importante citar los de Winkler y Smith (7) y Speybroeck et al (8). Una revisión acerca del uso de los métodos bayesianos en la evaluación de los resultados de ensayos clínicos, meta análisis, datos espaciales, modelos de supervivencia, estudios de validación de pruebas para diagnóstico clínico, genética molecular, modelamiento espacio temporal y toma de decisiones clínicas durante un período de 25 años (1982-2007) fue realizada por la profesora Deborah Ashby (9). En esa revisión la autora concluye que cada vez son más los estadísticos y bioestadísticos que usan las aproximaciones bayesianas para obtener sus resultados. En este artículo se presenta el proceso de análisis estadístico de los datos producidos en un estudio de validación de una prueba para diagnóstico clínico mediante métodos bayesianos comparando los resultados con los obtenidos utilizando la forma convencional de análisis.

Materiales y métodos.

Mogollón et al (10) realizaron un estudio para establecer las características de desempeño diagnóstico de la Colangiopancreatografía magnética (CRM) temprana en pacientes con pancreatitis Aguda de origen Biliar (PAB) para la detección de coledocolitiasis y usando como prueba de referencia la Colangiopancreatografía endoscópica retrógrada (CPRE). En su estudio, los autores tomaron datos de 154 historias clínicas pertenecientes a individuos con pancreatitis aguda de origen biliar leve que fueron expuestos a CRM. En 153 de los casos se contaba con el resultado de la ecografía como estudio adicional a la CRM y la CPRE, así que se construyó un procedimiento diagnóstico adicional con la combinación en paralelo de los resultados de la CRM y le Ecografía y se asumió que el resultado del tamizaje es positivo cuando al menos una de las dos pruebas lo es, lo que implica que un resultado negativo solo es posible si ambas pruebas arrojaban resultados negativos. Al momento de procesar los datos, se encontró que 86 sujetos contaban con resultado de CPRE y CRM, 81 tenían resultado de Ecografía y CPRE y 80 tenían datos de las tres pruebas, así que estos tres valores fueron los tamaños de muestra finales para el análisis de los datos.

Modelo estadístico asociado a la situación de estudio: caso continuo

Se puede entender el proceso de aplicación de las pruebas para diagnóstico como un experimento estadístico en el que se observan cuatro eventos (sucesos) de interés:

$A = \{\text{el resultado de la CPRE es positivo para coledocolitiasis}\}$
 $A^1 = \{\text{el resultado de la CPRE es negativo para coledocolitiasis}\}$
 $B = \{\text{el resultado de la CRM es positivo para coledocolitiasis}\}$
 $B^1 = \{\text{el resultado de la CRM es negativo para coledocolitiasis}\}$

A la ocurrencia de cada uno de los citados eventos podemos asociarle una variable aleatoria que permita presentar el resultado de manera cuantitativa de la siguiente forma: Sean las variables E y Y tales que:

$$E = \begin{cases} 1 & \text{si ocurre el evento } A \\ 0 & \text{si ocurre el evento } A^1 \end{cases} \quad Y = \begin{cases} 1 & \text{si ocurre el evento } B \\ 0 & \text{si ocurre el evento } B^1 \end{cases}$$

Definimos entonces la capacidad de la CRM detectar un caso de coledocolitiasis entre los individuos que tienen CPRE positivo como la sensibilidad, la cual, en términos del enfoque frecuentista de la estadística, la asumimos como un valor fijo denominado parámetro que requiere ser estimado a partir de la información brindada por los datos recolectados y que en términos bayesianos definiremos como cantidad aleatoria de interés de la que queremos conocer su distribución de probabilidad (distribución a posteriori) y la cual será obtenida al actualizar la probabilidad inicial (a priori), establecida a partir de la información de un especialista en el tema (proceso de elicitation) con la información contenida en los datos de la muestra de historias clínicas (verosimilitud). De igual forma se definirá la Especificidad o capacidad de la CRM de detectar individuos con CPRE negativa. En términos más formales, la sensibilidad y la especificidad son ambas probabilidades condicionales que reflejan el desempeño en campo de la CRM. Con el fin de abreviar y presentar el modelo estadístico que interpreta la situación real se denotará la sensibilidad con la letra griega α , la especificidad con la letra griega β y la prevalencia o probabilidad de encontrar una persona enferma (infectada) en la población general como π de modo que:

$$\begin{aligned} \tau &= \text{Prob (CPRE positiva)} = \text{Prob (A)} \\ \alpha &= \text{Prob (CRM positiva dado que CPRE positiva)} = \text{Prob (B|A)} \\ \beta &= \text{Prob (CRM negativa dado que CPRE negativa)} = \text{Prob (B^1|A^1)} \end{aligned}$$

El término “dado que” (|) indica que las probabilidades buscadas son “condicionales” es decir, necesitamos estimar la proporción de individuos con CRM positiva dentro de la población de individuos con CPRE positiva (en el caso de la sensibilidad) y de igual forma, estimar la proporción de individuos con CRM negativa dentro de la población de personas con CPRE negativa. En términos de las variables aleatorias definidas antes tenemos que: $\alpha = P(Y=1 | E=1)$ $\beta = P(Y=0 | E=0)$. En este caso, ya contamos con alguna información básica acerca de α y β como por ejemplo que ambas variables aleatorias solo pueden tomar valores no negativos en el intervalo (0,1) y que ambas son proporciones de éxitos dentro de sus respectivas poblaciones (sensibilidad dentro de los CPRE positivos y especificidad dentro de los CPRE negativo) lo cual implica que las mismas pueden ser asumidas como parámetros que indexan distribuciones de probabilidad que pueden describir el comportamiento de los datos. Con base en lo expuesto, podemos decir que para una sola repetición del experimento estadístico, el comportamiento natural de la variable aleatoria Y (resultado de la CRM) dentro del grupo de individuos con CPRE positiva, puede ser modelado con una distribución de probabilidad Bernoulli indexada por el parámetro α de modo que la suma de individuos con resultado

positivo dentro de la muestra de estudio, es una realización de una variable aleatoria cuyo comportamiento natural puede ser modelado a través de una distribución binomial de probabilidades indexada por el mismo parámetro. Haciendo la analogía para el grupo de individuos con CPRE negativa, se tiene que el evento éxito en ese caso es tener una CRM negativa, lo cual implica que Y se distribuye Bernoulli con parámetro (β) y la suma se distribuirá Binomial con éste parámetro.

Verosimilitud.

Para realizar el procedimiento de estimación, se requiere de explicitar las distribuciones de probabilidad asignadas al comportamiento de los datos y establecer una forma para representar la distribución que mejor describe el conocimiento del experto o la información encontrada en fuentes externas al experimento. Definimos las variables aleatorias A y D tales que $A = \sum_{i=1}^{r_1} y_i$ para todos los $y_i=1$ dentro del grupo de individuos con CPRE positiva y $D = \sum_{i=1}^{r_2} y_i$ para todos los $y_i=0$ en el grupo de individuos con CPRE negativa. Como se dijo antes, el comportamiento natural de A y D se puede modelar en términos de probabilidades usando una distribución Binomial de modo que:

$$P(A=a|\alpha) = \binom{r_1}{a} \alpha^a (1-\alpha)^{r_1-a} \quad y \quad P(D=d|\beta) = \binom{r_2}{d} \beta^d (1-\beta)^{r_2-d}$$

Donde r_1 y r_2 son las cantidades de sujetos que tienen CPRE positiva y CPRE negativa respectivamente lo cual implica que el total de sujetos evaluados en el estudio será $n = r_1 + r_2$.

Dado que el interés principal radica en estimar α y β de los cuales se desconoce su forma probabilística de comportamiento pero se sabe que toman valores en un continuo (no existen saltos entre un valor y otro) en el intervalo que va de cero a uno, se requiere expresar la información que está contenida en los datos de una forma que la misma pueda “conversar” en forma probabilística con la forma probabilística de la información externa obtenida a través de un especialista. En vista de que, ya se tienen los datos y es posible verlos, lo que se desconoce son las cantidades aleatorias α y β , entonces, se puede escribir la estructura del modelo de probabilidad en forma de verosimilitud, es decir como una expresión de la información que tienen los datos acerca de las cantidades desconocidas, por tanto:

$$L(\alpha|a) = \binom{r_1}{a} \alpha^a (1-\alpha)^{r_1-a} \quad y \quad L(\beta|d) = \binom{r_2}{d} \beta^d (1-\beta)^{r_2-d}$$

Distribuciones a priori

El proceso de tomar la información que tiene el especialista y expresarla en forma de una distribución de probabilidades, se ha denominado proceso de elicitación de la distribución a priori el cual es un proceso complejo con el que se formalizan las creencias que tiene un individuo acerca del comportamiento natural de un fenómeno que se ven influenciadas por su conocimiento de la evidencia obtenida en investigaciones previas sobre el asunto, su experiencia clínica o simplemente su grado de exposición al evento (11). La elicitación de información a priori ha sido tema de interés de investigadores tanto de la estadística como de la psicología. Tres trabajos de revisión bastante completos que abordan las heurísticas, los sesgos y una aproximación a la explicación de los procesos cognitivos y psicológicos implicados en un proceso de elicitación son los de Hogart (12), Garthwaite et al. (13) y Alonso (14). Desde el punto de vista más estadístico del asunto, autores como Kadane et al (15,16), Chaloner y Duncan (17) y Tovar (18) han concentrado los esfuerzos en desarrollar métodos que permitan objetivar la forma de la distribución a priori.

De acuerdo con Spiegelhalter et al (19) no es necesario seleccionar una única distribución a priori para hacer el análisis y este mismo autor recomienda la presencia de una “comunidad” de distribuciones a priori que podrían ser usadas, como las a priori de referencia que son distribuciones que representan la mínima cantidad posible de información a priori disponible, generalmente útiles cuando no se cuenta con un especialista en el tema o se tiene tan poca información publicada sobre el comportamiento natural de las cantidades aleatorias de interés que consideramos que la misma no es relevante, generalmente, las a priori de referencia se utilizan para tener una línea de base con la cual comparar otras distribuciones o los resultados obtenidos usando métodos clásicos. Existen las distribuciones a priori clínicas (clinical priors) que formalizan la opinión de una persona bien informada o especializada en el fenómeno de interés. Otro grupo de distribuciones a priori son las escépticas, las cuales formalizan creencias poco positivas en el comportamiento de las cantidades respecto a las hipótesis o supuestos que soportan el estudio y finalmente, están las a priori entusiastas que expresan la opinión del especialista que confía en que los resultados del estudio van a ser los definidos a priori en el diseño. Para el caso de la estimación de la sensibilidad y la especificidad del CRM, se procederá usando tres formas de distribuciones a priori, en la primera se asumirá el conocido principio de La Place, de que todos los posibles valores de α y β en el intervalo (0,1) tienen la misma probabilidad de ocurrir, es decir no se está privilegiando ningún valor o conjunto de valores en especial, esta forma de distribución a priori se formaliza utilizando una distribución de probabilidad Uniforme en el intervalo (0,1). Otra distribución que se tomará en consideración será la obtenida a través de los resultados obtenidos en estudios de validación de la CRM en otros países o datos tomados de la literatura y la tercera forma será la obtenida a través de la información brindada por un médico especialista en cirugía de vías biliares. Con base en el conocimiento básico que se tiene del comportamiento de α y β como variables aleatorias, se debe buscar una distribución de probabilidades que modele el comportamiento natural de variables continuas en un intervalo acotado (en este caso entre cero y uno) y una buena candidata para tal fin es la distribución Beta de probabilidades la cual tiene la forma analítica: $f(x|s,t) = kx^{s-1}(1-x)^{t-1}$ donde k es una constante que hace que la integral de la función evaluada en todos los posibles valores de x sea igual a uno y s, t son los parámetros que indexan la distribución los cuales son conocidos como hiperparámetros pues son los parámetros de la distribución a priori. La Beta es considerada también una distribución a priori conjugada para una verosimilitud con comportamiento binomial, es decir, si usamos una a priori Beta y la combinamos mediante el teorema de Bayes con una verosimilitud binomial, obtenemos como distribución de probabilidades a posteriori otra distribución Beta, lo que en términos prácticos facilita bastante el proceso matemático de estimación.

Se tiene entonces el trabajo de obtener los valores de s y t que se asocian a cada una de las distribuciones Beta que serán usadas como distribución a priori. En el primero de los casos, cuando se asume que todos los valores de α como todos los valores de β tienen la misma probabilidad de ser observados, es decir, se asume que las cantidades aleatorias α y β se distribuyen de manera uniforme en el intervalo (0,1), es exactamente lo mismo que ajustar como a priori una Beta con hiperparámetros $s=1$ y $t=1$. Para el segundo y tercer caso, es necesario construir intervalos de valores dentro de los cuales se espera tener una mayor probabilidad de encontrar los valores de α y β y luego aplicar el procedimiento desarrollado por Tovar (18) para encontrar el valor exacto de los hiperparámetros. En términos generales, el citado procedimiento toma un intervalo de valores en donde se puede encontrar con mayor probabilidad la cantidad aleatoria de interés, el

cual es obtenido a través de conversaciones con el especialista o de la revisión de la literatura. Se toma el valor medio del intervalo como el valor esperado para la cantidad aleatoria y se iguala a la forma matemática de la esperanza de una distribución Beta. Luego utilizando el teorema de Chebychev que es una desigualdad bastante conocida en la teoría de la probabilidad, se obtiene una aproximación de la varianza de la cantidad aleatoria dentro del intervalo elicitado y con estos dos valores se ajusta un sistema de ecuaciones y se obtienen los valores de s y t para la distribución Beta a priori. Otro método que puede ser utilizado es el del Joseph et al (20) el cual parte de un intervalo fijo pero su aproximación de desviación estándar de la cantidad aleatoria la hace usando el primer cuartil del intervalo.

Para construir el intervalo a priori donde se espera con una alta probabilidad encontrar el valor de la cantidad de interés usando información publicada, se revisaron los artículos de Pisani et al (21) y Rocha et al (22). Esos estudios reportaron valores exactos de la sensibilidad y la especificidad sin intervalos de confianza, así que se construyeron intervalos tomando los valores reportados por los dos estudios para cada parámetro de desempeño del CRM. Se consultó también a un especialista en cirugía de vías biliares, quien informó que la sensibilidad de la CRM debería tomar valores en el intervalo 0.8 – 0.9 y podía ser mayor si se acompaña con un valor positivo de la fosfatasa alcalina y que para la especificidad, el valor podía estar entre 0.85 y 0.90 cuando se cuenta con un radiólogo que tenga experiencia evaluando colangiogramas. Para cada uno de los intervalos se obtuvo el valor de los hiperparámetros utilizando el procedimiento de Tovar (18). Ver Cuadro 1.

Distribuciones a posteriori: una vez conocidos los hiperparámetros de las distribuciones de probabilidad que describen (antes de observar los datos) el comportamiento de la cantidad aleatoria de interés, se procede a obtener la forma analítica y gráfica de la distribución de probabilidades actualizada o modificada por la información contenida en los datos, la cual es denominada distribución a posteriori. El proceso de actualización se realiza utilizando el teorema de Bayes de la siguiente forma:

$$f(\alpha|a, s_1, t_1) = \frac{L(\alpha|a) f(\alpha)}{\int_0^1 L(\alpha|a) f(\alpha) d\alpha} \quad f(\beta|d, s_2, t_2) = \frac{L(\beta|d) f(\beta)}{\int_0^1 L(\beta|d) f(\beta) d\beta}$$

Para el caso tratado en este artículo, $f(\alpha)$ y $f(\beta)$ son distribuciones (a priori) Beta (s_i, t_i) $i=1,2$ de modo que la distribución a posteriori obtenida después de algunos cálculos matemáticos sencillos, es una nueva distribución Beta (v_1, v_2) con $v_1 = s_1 + a$, $v_1 = t_1 + (r_1 - a)$ y $v_2 = s_2 + d$, $v_2 = t_2 + (r_2 - d)$. La familia de distribuciones Beta tiene la ventaja de ser sencilla de manejar y el hecho de estar implementada en la mayoría de los programas estadísticos para computador la hace muy útil y práctica para trabajar. Los valores para los parámetros de las distribuciones a posteriori aparecen en el Cuadro 1.

Cuadro 1. Intervalos propuestos para la construcción de distribuciones a priori y valores obtenidos para los parámetros para las distribuciones a priori y a posteriori

Cantidad Aleatoria de interés	Intervalos de información a priori	Hiperparámetros de la distribución a priori		Parámetros de la distribución Beta a posteriori**	
		S	t	u	v
Sensibilidad (α)	0,00 – 1,00	1,0	1,0	69 56	3 17
	0,86 - 0,95	767,6	80,6	835,6 822,6	82,6 96,6
	0,80 – 0,90*	866,0	152,9	934,0 921,0	154,9 168,9
Especificidad (β)	0,00– 1,00	1,0	1,0	8,0 4,0	10,0 7,0
	0,67 – 1,00	83,7	16,5	90,7 86,7	25,5 22,5
	0,85 – 0,90*	48,9	6,99	55,9 51,9	16,0 12,9

*Valores obtenidos a partir de la información brindada por un especialista

**El valor superior es para la distribución de la CRM y el inferior para la combinación CRM+ECOGRAFIA

Aun cuando, toda la información sobre la cantidad aleatoria de interés, se concentra en la distribución de probabilidad a posteriori que describe el comportamiento de la misma después de combinar la información externa contenida en la cabeza del especialista (subjectiva) o en documentos o archivos históricos, con la información brindada por una muestra de mediciones obtenidas en campo o experimentalmente, también es posible resumir dicha información en medidas índice de la misma forma que se hace en la estadística clásica. La medida índice más comúnmente utilizada es la esperanza a posteriori, es decir, la media de la distribución a posteriori, sin embargo, dependiendo de la forma que presente el comportamiento posterior de los valores de la cantidad aleatoria (puede ser pronunciadamente asimétrica o tener algo o mucha simetría) se puede decidir por usar la moda o la mediana de la distribución. Para el caso aquí presentado, se podría resumir la información a posteriori sobre α y β usando la media (valor esperado) o la moda (el valor con mayor densidad a posteriori), las cuales tienen las siguientes fórmulas de cálculo:

$$\hat{\alpha} = E(\alpha|a) = \frac{u_1}{u_1 + v_1}, \quad \hat{\beta} = E(\beta|d) = \frac{u_2}{u_2 + v_2} \quad \text{para las medias y} \quad \tilde{\alpha} = \frac{u_1 - 1}{u_1 + v_1 - 2}, \quad \tilde{\beta} = \frac{u_2 - 1}{u_2 + v_2 - 2}, \quad \text{en caso de tomar la moda. Esta última fórmula puede}$$

ser aplicada siempre y cuando ambos parámetros de la distribución a posteriori sean mayores que uno. Desde el punto de vista de la inferencia estadística clásica, las estimaciones de las cantidades desconocidas (parámetros) se presentan a través de una medida resumen (generalmente la media) acompañada de su intervalo de confianza, el cual es una forma de expresar la reproducibilidad del estudio ya que brinda información acerca de los valores que toma el estimador usado en un porcentaje de veces (la confianza) que se repitiera la experiencia de obtención de datos. Desde el punto de vista Bayesiano, se reporta la región de credibilidad del (1-γ)%, cuya interpretación es totalmente diferente a la del intervalo de confianza ya que, la región de credibilidad es tomada directamente de la distribución a posteriori de la cantidad aleatoria, por lo tanto, es el rango de valores dentro del cual está el (1-γ)%, de los valores de dicha cantidad, por lo tanto (1-γ) es una probabilidad y no una confianza.

Caso discreto este procedimiento es el que generalmente se utiliza para estimar los valores predictivos positivo y negativo de un nuevo test en evaluación. Utilizando la notación con la que se definió el modelo estadístico asociado a la situación, el valor predictivo positivo (VPP) se puede obtener aplicando la siguiente fórmula: $VPP = P(A|B) = \frac{P(A \cap B)}{P(B)}$ entonces, $VPP = P(E = 1|Y=1) = \frac{P(E = 1, Y=1)}{P(Y=1)}$ aquí es importante destacar que a diferencia de la sensibilidad y la especificidad que también son probabilidades condicionales, los valores predictivos generalmente no son estimables a partir de los datos obtenidos en campo como si ocurre con las características de la prueba en evaluación. El VPP y el VPN como su nombre lo dice, 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 Gold Standard, situación común cuando se aplica la prueba tamiz de manera masiva después de ser validada y cuya estimación puede ser directamente obtenida aplicando la fórmula característica del teorema de Bayes a los datos del estudio de la siguiente forma:

$$VPP = P(E = 1|Y=1) = \frac{P(E = 1, Y=1)}{P(Y=1)} = \frac{P(E = 1) P(Y=1|E=1)}{P(E = 1) P(Y=1|E=1) + P(E = 0) P(Y=1|E=0)} = \frac{\tau\alpha}{\tau\alpha + (1 - \tau)(1 - \beta)}$$

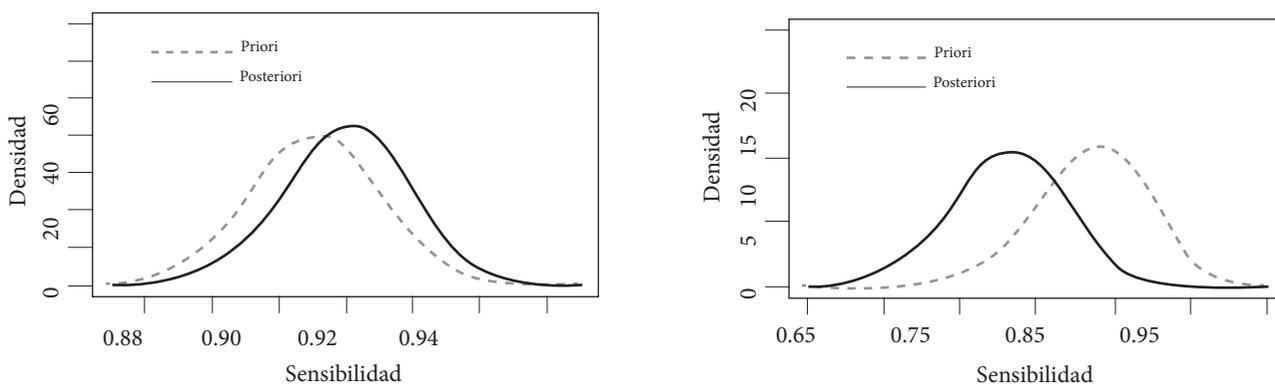
Es claro entonces, que el VPP depende de la proporción de personas con el evento de interés en la población general (prevalencia de enfermos/infectados) y las características de desempeño del procedimiento diagnóstico. Un análisis similar puede ser realizado para el VPN.

En muchos estudios, el diseño tiene alguna característica limitante o 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 del estudio 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 la cual está resumida en la sensibilidad cuando se trata del VPP y por la especificidad cuando se desea estimar el VPN. En el caso discreto, el VPP y el VPN son valores de probabilidad puesto que se está estimando a partir de valores puntuales de las características de la prueba en evaluación, caso contrario a lo que ocurre en el caso continuo donde se obtiene la forma de la distribución predictiva de probabilidades con la que se pueden obtener la probabilidad de ocurrencia de algún conjunto de valores que pueden ocurrir al hacer repeticiones del experimento de validación en campo. El artículo de Winkler and Smith (7) hace un estudio detallado acerca del VPP y el VPN.

Resultados

Con las distribuciones a priori elicítadas, se obtuvieron las distribuciones a posteriori de las cantidades aleatorias de interés, en este caso, los índices de desempeño diagnóstico tanto de la CRM como de la combinación CRM y Ecografía. Los gráficos obtenidos para la prueba tamiz resultante de la combinación CRM mas Ecografía son bastante similares a las de la Figura 1 así que se omite su presentación.

De acuerdo con la Figura 1, se puede observar que el comportamiento posterior de los índices se puede asumir simétrico dentro de un intervalo, sin embargo, para cada caso se calcularon la media, la mediana y la moda de la distribución para observar posibles distanciamientos de la citada característica observándose que en todos los casos los valores de las medidas de tendencia central son bastante similares. Se tomó entonces, la media como el valor indicador que resume la información de la cantidad aleatoria de interés. Ver Cuadro 2



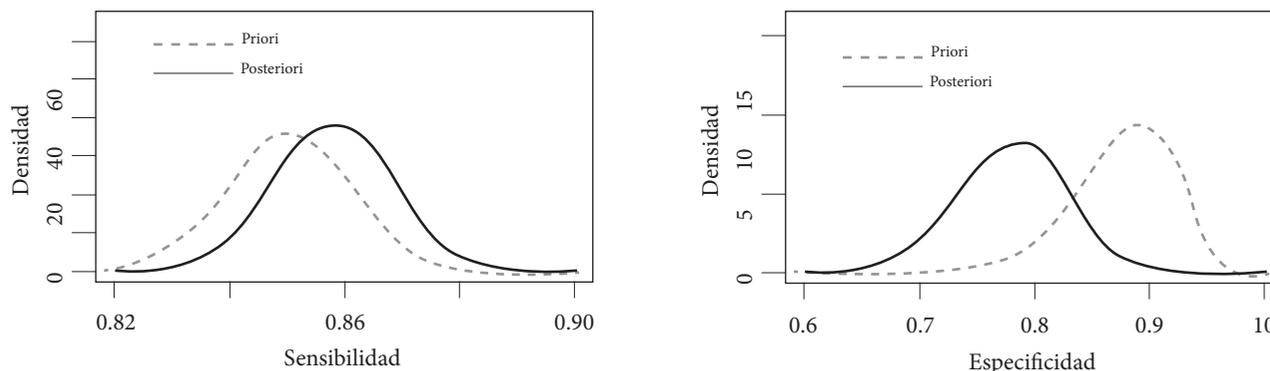


Figura 1. Formas de las distribuciones a priori y a posteriori para la CRM obtenidas a partir de información publicada (parte superior) e información suministrada por un médico cirujano (parte inferior).

Cuadro 2. Estimaciones de los parámetros de desempeño de las diferentes pruebas para diagnóstico clínico usando CPRE como prueba de referencia.

Método	Prueba	Sensibilidad	Especificidad
Inferencia Clásica* (método de máxima verosimilitud)	CRM	97.1 (92.5-100)	43.8 (16.3-71.2)
	Ecografía+CRM	77.5 (67.0-87.9)	33.3 (0-69.7)
Bayes usando a priori de referencia**	CRM	95.8 (90.2-99.1)	44.0 (22.9-67.0)
	Ecografía+CRM	76.7 (66.4-85.6)	36.4(12.1-65.2)
Bayes**(a priori obtenida con información publicada)	CRM	91.0 (89.0-93.0)	78.0 (70.1-85.0)
	Ecografía+CRM	89.5 (87.4-91.4)	79.4 (71.3-86.4)
Bayes** (a priori obtenida a través de un médico especialista)	CRM	85.8 (83.6-87.8)	77.8 (67.5-86.5)
	Ecografía+CRM	84.5 (82.3-86.6)	80.0(69.5-88.7)

*Se presenta el intervalo de confianza del 95% junto a la estimación del indicador de desempeño

**Se presenta la región de credibilidad del 95% de probabilidad junto a la estimación del indicador de desempeño

Cuando se utiliza una distribución a priori de referencia, las estimaciones no son muy diferentes a las observadas utilizando el método de verosimilitud máxima (método clásico) lo cual es de esperarse dado que a priori se asume que todos los posibles valores del intervalo (0-1) tienen la misma probabilidad de ser considerados para el proceso de estimación. Algunos autores usan el término de a priori no informativa para este caso y siempre se espera que el peso de la información suministrada por los datos sea mayor en la distribución a posteriori. Sin embargo, es importante resaltar que en este estudio, las regiones de credibilidad del 95% tienden a tomar un rango de valores más estrecho que el observado en los clásicos intervalos de confianza. Si se utiliza como distribución a priori la información contenida en artículos publicados por investigadores que han realizado estudios similares en otros países, la estimación de la sensibilidad de la CRM es algo menor que la obtenida usando máxima verosimilitud, pero la de la especificidad es mayor indicando que la información colocada puede ser bastante optimista o los datos colectados presentan limitaciones para estimar esta característica. Los resultados con este tipo de a priori no varían mucho cuando se toma solo la CRM con respecto a la combinación de resultados de la misma con la ecografía. En los resultados asociados a la distribución a priori obtenida a través de un experto, se observa una leve disminución en la estimación de la sensibilidad respecto a la obtenida con la anterior distribución a priori. En el caso de la especificidad, el comportamiento es similar. Ver Cuadro 2.

Dado que se utilizaron tres distribuciones a priori diferentes para obtener las estimaciones de la sensibilidad y la especificidad de la CRM y la prueba combinada, se hace necesario establecer cuál

de los mismos es el más adecuado para presentar los resultados. Plantear diferentes modelos que expliquen el comportamiento natural del fenómeno en estudio, es un procedimiento común entre los estadísticos tanto clásicos como bayesianos, pues no es posible concluir que una sola aproximación es la más adecuada para explicar una situación problemática cuando pueden existir una gran cantidad de alternativas dependiendo de las heurísticas usadas para entender el problema y las características propias del mismo. Existen muchos métodos de selección de modelos (alternativas de análisis) reportados en la literatura y estudiados por los diferentes autores, desde el punto de vista bayesiano, es bastante común utilizar los factores de Bayes (BF por su nombre en inglés) como método de selección. El BF utiliza el cociente entre las probabilidades predictivas a posteriori de observar una nueva muestra del mismo tamaño con la cantidad de éxitos observado en el estudio, asumiendo que el valor de la cantidad aleatoria es el obtenido en la estimación para cada modelo en comparación. Es decir, se sabe que las cantidades de individuos con resultado de CRM positivo es un número de éxitos dentro de la muestra de pacientes con resultado positivo para la CPRE y lo que se estima es la probabilidad de éxito que es la misma proporción de éxitos en la población de individuos con resultado positivo en la CPRE. La estimación de dicha probabilidad es precisamente la esperanza (media) de la distribución a posteriori para la cantidad aleatoria. Para el caso de este artículo, el interés se centra en evaluar cuál de las dos distribuciones a priori elicítadas es mejor opción para la estimación de la proporción de interés, si se sabe que para ambas distribuciones a priori se obtuvo una a posteriori combinando las mismas mediciones de la muestra de estudio, entonces las distribuciones a posteriori solo cambian por la a priori usada ya que la verosimilitud es constante. En

este caso, se calculó la probabilidad de obtener el número observado de éxitos dentro de las muestras asumiendo que el verdadero valor de la probabilidad de éxito era el obtenido con cada uno de los modelos en competencia, es decir, se calculó la probabilidad de tener 68 resultados positivos para la CRM entre los 70 individuos con CPRE positiva asumiendo primero que la probabilidad de tener CRM positiva en la población es 0,91 (modelo obtenido con a priori de la literatura) y luego asumiendo que la probabilidad de tener CRM positiva es 0.858 (modelo obtenido usando información de una especialista). El cálculo de las probabilidades predictivas para el caso ilustrativo se realizó de la siguiente manera:

$$P(A = 68 | \alpha_1 = 0.91) = \binom{70}{68} 0.91^{68} (1 - 0.91)^2 = 0.0321$$

$$P(A = 68 | \alpha_2 = 0.858) = \binom{70}{68} 0.858^{68} (1 - 0.858)^2 = 0.0015$$

Con lo que se tiene que el BF es 21.3 indicando que la fuerza de la evidencia es mayor a favor de que los datos hayan sido generados por un proceso binomial cuya probabilidad de éxito es 0,91.

Con lo que se tiene que el BF es 21.4 indicando que la fuerza de la evidencia es mayor a favor de que los datos hayan sido generados por un proceso binomial cuya probabilidad de éxito es 0.91. Al hacer el mismo análisis para la especificidad, el factor de Bayes es 1.04 indicando que ninguno de los dos modelos presenta mayor evidencia a su favor. De manera similar se procedió con los modelos obtenidos para la especificidad de la CRM y para las dos cantidades aleatorias en el diagnóstico combinado CRM+ECOGRAFIA. Ver Cuadro 3.

Cuadro 3. Probabilidades predictivas y factores de Bayes (M1/M2) para los modelos utilizados en el proceso de estimación

Prueba	Indicador	A priori literatura (M1)	A priori especialista (M2)
CRM	Sensibilidad	0.0321	0.0015
	Especificidad	0.0024	0.0026
	Sensibilidad	0.0016	0.0337
CRM + Ecografía	Especificidad	0.0032	0.0028
	Factor de Bayes	21.96	
CRM	Especificidad	0.939	
	Factor de Bayes	0.046*	
Combinación	Especificidad	1.167	

*BF=21.7 para M2/M1

De acuerdo con la escala de interpretación del BF establecida por Jeffreys, en todos los casos se puede decir que se tiene muy escasa evidencia a favor del modelo M1 comparándolo con el modelo M2, lo cual implica que podría tomarse cualquiera de los dos modelos para hacer las estimaciones. Apoyándose en argumentos heurísticos, se podrían tomar los resultados del modelo M2 para dar información sobre el problema de interés.

Conclusiones

La inferencia bayesiana ha venido tomando cada vez más fuerza en el análisis de datos producto de proyectos de investigación en el campo de los estudios médicos. Este tipo de análisis de datos es más natural y acorde con la forma como razonan los seres humanos a la

hora de tomar decisiones, es decir, se tiene una información que es almacenada durante el tiempo de exposición al fenómeno de interés la cual es actualizada con los información observada en la naturaleza y a partir de la fusión se obtienen resultados que permiten llegar a conclusiones y tomar decisiones. El enfoque bayesiano, puede ser bastante útil cuando, como en la situación abordada en este estudio, se cuenta con un tamaño de muestra que puede considerarse “pequeño” para lograr el objetivo del estudio teniendo un alto poder en la obtención de los resultados. El libro de Sullivan (23) puede ser una buena referencia acerca de los diseños y los cálculos de tamaño de muestra para estudios de evaluación de desempeño en pruebas diagnósticas. En estudios de este tipo, que cuentan con una muestra “pequeña” de individuos o historias clínicas como es el caso tratado en este artículo, se corre el riesgo de que alguno de los grupos formados por el resultado de la prueba usada como gold standard o prueba de referencia se encuentren muy pocos resultados éxito lo que conlleva a que se sobreestime alguna o las dos cantidades de interés a ser estimadas, efecto que se va a reflejar en los intervalos de confianza obtenidos los cuales tenderán a ser más anchos de lo esperado y por ende poco informativos. En situaciones de esta clase, es muy útil estimar mediante métodos Bayesianos, los cuales adicionan a la probablemente poca información contenida en los datos, la información sobre el fenómeno contenida en fuentes externas. De acuerdo con la revisión de la literatura realizada para obtener información a priori sobre el desempeño de la CRM en conjunto con la Ecografía, no se encontraron estudios en los que se haya utilizado la combinación de resultados para realizar el diagnóstico, aunque se encontraron estudios en los que se uso solo la CRM. Este hecho podría afectar los resultados para la combinación ya que el análisis bayesiano se realizó una distribución a priori construida a partir de la CRM, lo cual no es totalmente realista para la combinación. Los resultados muestran que hay un aumento en la estimación de la sensibilidad y la especificidad cuando se asume la información a priori de la CRM, lo cual tiene sentido ya que lo esperado es que el diagnóstico combinado tenga un desempeño similar y no peor al de la CRM sola. Se puede decir entonces, que si se asumen los resultados de Mogollon et al(10) como una primera aproximación al estudio de los parámetros de desempeño de la combinación, en este trabajo, los mismos son mejorados buscando tener una aproximación más precisa. Aun cuando, se evaluaron dos distribuciones a priori “informativas” una obtenida a partir de información publicada y otra a partir de un especialista, en el momento obtener las estimaciones, se observó que para la sensibilidad, el criterio de selección de modelos permite concluir que el modelo con a distribución priori elicitada usando datos de la literatura podría presentar un mejor comportamiento para la estimación de la cantidad desconocida, mientras que para la estimación de la especificidad, el modelo a ser seleccionado sería el ajustado usando la distribución a priori elicitada a través del experto. Sin embargo, el resultado para la sensibilidad no es del todo concluyente ya que el BF toma un valor bastante próximo de la unidad indicando que la diferencia entre los dos modelos no es realmente importante, lo cual se explica al observar con cuidado los intervalos usados para elicitar las distribuciones que son muy similares entre ellos. En la estimación de la especificidad, se observa un poco más de diferencia entre los intervalos en los que se espera encontrar a priori el valor de la cantidad desconocida y dicha diferencia se refleja en el comportamiento de los modelos y el BF. En este caso, el modelo que podría ser seleccionado es el obtenido a partir de los datos proporcionados por el experto, lo que tiene es esperable puesto que el intervalo a priori es más estrecho que el intervalo obtenido a través de los datos publicados. En este estudio se construyeron intervalos a priori para elicitar las distribuciones a priori utilizando información publicada en otros países lo cual es viable en situaciones en las que

es difícil contar con la participación de un especialista en el tema, sin embargo, es importante ser consciente de las implicaciones que puede tener usar este tipo de información ya que se pueden presentar variaciones entre los estudios debidas a características propias de los individuos y las condiciones empleadas en el diseño y planeación del estudio en el país específico que no sean totalmente comparables con las condiciones propias del otro país. Siempre que sea posible contar con información obtenida a partir de especialistas que cumplan con algunos criterios deseables de exposición y conocimiento del tema es mejor considerar esta opción para elicitar las distribuciones a priori. Finalmente, además del BF, existen muchos otros procedimientos para realizar la selección del modelo que mejor explique una situación real observada en la naturaleza (AIC, BIC, DIC, etc) cuyo nivel de complejidad estadístico matemática no están dentro del alcance del presente artículo, razón por la que no son abordados por el autor.

Agradecimientos

El autor expresa sus más sinceros agradecimientos a la doctora Geanny Mogollón por autorizar el uso de los datos publicados en la Revista Cubana de Cirugía para ilustrar los conceptos tratados en este trabajo y a los estudiantes Diosimar Cardoza y Víctor Fabián Chávez por su valioso apoyo en la parte computacional y de cálculo.

Referencias

- Siva L.C., Benavides A. Apuntes sobre subjetividad y estadística en la investigación en salud. *Rev. Cubana Salud Pública*, 2003; 29(2): 170-73
- Gurrin L.C., Kurinczuk J.J., Burton P.R. Bayesian statistics in medical research: an intuitive alternative to conventional analysis. *J Eval. Clin. Pract.* 2000; 6(2). 193-204
- Gill, C.J., Sabin L., Schmid C.H. Why clinicians are natural Bayesians. *BMJ*, 2005; 330:1080-83
- Silva, L.C., Benavides, A. El enfoque bayesiano: otra manera de inferir. *Gac. Sanit.* 2001; 15(4): 341-46
- Ashby D., Smith F.M. Evidence-based medicine as bayesian decision-making. *Stat. Med.* 2000, 16: 3291-3305
- Ashby, D. Bayesian methods. [cited junio 2014] in <http://www.wiley.com/legacy/wileychi/eob/bct/Cab001-.pdf> [Online]
- Winkler L., Smith J. On uncertainty in Medical Testing. *Med. Decis. Making.* 2004; 24: 654-58.
- Speybroeck N., Devleeschauwer B., Lawrence J., Berkvens D. Misclassification errors in prevalence estimation: Bayesian handling with care. *Int. J. Public Health.* 2013;58(5): 791-5
- Ashby D. Bayesian Statistics in medicine: A 25 year review. *Stat. Med.* 2006; 25(21): 3589-631
- Mogollon G., Sefair C., Upegui D., Tovar J.R. Colangiopancreatografía magnética: valor diagnóstico para detectar coledocolitiasis en pacientes con pancreatitis aguda leve. Online <http://www.revquirugia.sld.cu/index.php/cir/article/view/68/44>
- Johnson S.R., Tomlinson G.A., Hawker G.A., Granton J.T., Feldman B.M. Methods to elicit beliefs for Bayesian priors: a systematic review. *J. Clin. Epidemiol.* 2010; 63:355-69
- Hogarth, R.M. Cognitive process and the assessment of subjective probability distributions. *J. Am. Stat. Assoc.* 1975; 70(14): 271-89
- Garthwaite, P.H., Kadane, J.B., O'Hagan A. Statistical methods for eliciting probability distributions. *J. Am. Stat. Assoc.* 2005; 100(470): 680-700
- Alonso D., Tubau E. Inferencias bayesianas: una revisión teórica. *Anuario de Psicología.* 2002; 33(1); 25-47
- Kadane, J.B. Wolfson, L.J. Experiencies in elicitation. *The Statistician* 1998; 47: 2-18.
- Kadane, J.B., Dickey, J.M., Winkler, R.L., Smith, W. and Peters, S.C. Interactive elicitation of opinion for a normal linear model. *J. Am. Stat. Assoc.* 1980; 75: 845-854.
- Chaloner, K.M. Duncan, G.T. Assessment of a Beta prior distribution: PM Elicitation. *The Statistician* 1983; 32: 174-180.
- Tovar J.R. Eliciting Beta prior distributions for binomial sampling. *Rev. Bras. Biom.* 2012; 30:159-72
- Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J. Royal Stat. Soc.* 1994; 157(3):357- 416
- Joseph L., Gyorkos T.W., Coupal L. Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard. *Am. J. Epidemiol.* 1995; 141 (3): 263-272.
- Pisani, J.C., Bacelar A., Malafaia O., Marcondes J., Gregori C.N., Nunes P.A. Estudo comparativo entre a colangiopancreatografia por ressonância magnética e a colangiopancreatografia endoscópica no diagnóstico das obstruções biliopancreáticas. *Arquivos de Gastroenterologia.* 2001; 38(3): 149-57
- M.S. Rocha, S.K. Ueda, M.C. Machado, H.W. Pinotti. Colangiopancreatografia por ressonância magnética: uma nova forma de avaliar as vias biliares e pancreáticas. *Rev Ass Med Brasil* 1998; 44(3): 226-8.
- Sullivan M. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford Statistical Science series 3, 2003