

Modelo de selección de variables Bayesiana aplicada
a la identificación de subgrupos en el análisis
coste–efectividad

Francisco José Vázquez–Polo
Universidad de Las Palmas de G.C., España
fjvpolo@dmc.ulpgc.es

Miguel Ángel Negrín
Universidad de Las Palmas de G.C., España
mnegrin@dmc.ulpgc.es

María Martel
Universidad de Las Palmas de G.C., España
mmartel@dmc.ulpgc.es

Elías Moreno
Universidad de Granada, España
emoreno@ugr.es

Francisco Javier Girón
Universidad de Málaga, España
fj_giron@uma.es

7 de marzo de 2014

Clasificación Código JEL: C11, C01, I1

Dirección de correspondencia: Francisco–José Vázquez–Polo, PhD, Universidad de Las Palmas de Gran Canaria, Departamento de Métodos Cuantitativos en Economía y Gestión, Campus de Tafira, 35017, Las Palmas de Gran Canaria, España; e–mail: fjvpolo@dmc.ulpgc.es.

Modelo de selección de variables Bayesiana aplicada a la identificación de subgrupos en el análisis coste–efectividad

Resumen

En el análisis coste–efectividad de tecnologías sanitarias es ampliamente reconocido que, en ocasiones, el tratamiento óptimo puede variar para distintos grupos de población, lo que se conoce en la literatura como presencia de subgrupos. Por tanto, definir con precisión los subgrupos debe ser una prioridad en el análisis. Recientemente, varios autores han propuesto una definición de subgrupo de pacientes sobre la base de las interacciones entre la variable indicadora del tratamiento y las características del paciente en el contexto de un modelo de regresión. En este trabajo proponemos el uso de un modelo de selección de variables Bayesiano para la identificación de subgrupos. Este nuevo enfoque proporciona resultados con una aplicabilidad más amplia y precisa. Un estudio de simulación asumiendo diferentes distribuciones para la eficacia y el costo y diferentes tamaños muestrales ha sido desarrollado para comparar ambas metodologías. El procedimiento de selección Bayesiano también es validado para un estudio con datos reales.

Palabras claves: Selección de variables Bayesiana, análisis coste–efectividad, modelos de regresión, distribución a priori intrínseca, análisis de subgrupos

1. Introducción

El análisis coste–efectividad puede interpretarse como un problema de decisión estadístico donde debemos elegir entre m tratamientos alternativos teniendo en cuenta tanto la efectividad como el coste del tratamiento. Se trata por tanto de un problema de optimización extensamente trabajado en la literatura de la economía de la salud (ver por ejemplo Barton *et al.*, 2008).

La decisión de qué tratamiento es óptimo se realiza a partir del beneficio neto $\mathbf{z} = (z_1, \dots, z_m)$, que combina coste y efectividad para cada tratamiento. El beneficio neto del i -ésimo tratamiento es definido como $z_i = R \cdot e_i - c_i$, donde c_i y e_i son el coste y la efectividad del tratamiento T_i , y R es la utilidad asignada a cada unidad de efectividad expresada en unidades monetarias. Se puede observar que, condicionada a la cantidad R , que es fijada por el decisor, la distribución del beneficio neto es determinada con la distribución bivalente de la efectividad y el coste.

Varios autores discuten que la función que define el comportamiento del coste y de la efectividad de un tratamiento $f_i(c_i, e_i|\theta_i, \mathbf{x})$, $i = 1, \dots, m$, donde $\theta_1, \dots, \theta_m$ son los parámetros a estimar, dependen también de un conjunto p de características del paciente o covariables $\mathbf{x} = (x_1, \dots, x_p)$. De esta forma

la decisión del tratamiento óptimo dependerá también de las características de los pacientes \mathbf{x} dando lugar a lo que la literatura en economía de la salud denomina subgrupos (Sculpher and Gafni, 2001; NICE, 2008; Espinoza *et al.*, 2011). El tratamiento óptimo puede ser distinto para cada subgrupo (Moreno *et al.*, 2012), por lo que uno de los objetivos prioritarios de investigación debe ser definir claramente que características de los pacientes generan subgrupos. En este trabajo proponemos resolver esta cuestión como un problema estadístico de selección de variables desde el punto de vista Bayesiano.

En la literatura reciente en economía de la salud, el análisis de subgrupos ha sido propuesto a través de la incorporación de interacciones entre las variables indicadoras de tratamiento y las características de la población susceptibles de generar subgrupos. La existencia de subgrupos se reduce a la realización de un test de hipótesis sobre la significación estadística del coeficiente de dicha interacción. Esta modelización adolece de algunas limitaciones, como por ejemplo el hecho de solo permitir la consideración de subgrupos para características poblacionales medidas a través de variables discretas (Stinnett and Mullahy, 1998; Pocock *et al.*, 2002; Willan *et al.*, 2004; Nixon and Thompson, 2005; Sculpher, 2008; Manca *et al.*, 2010; and Gomes *et al.*, 2012a,b, entre otros).

Entre los modelos propuestos por la literatura utilizaremos el modelo propuesto por Nixon and Thompson (2005) como modelo de referencia. La formulación de dicho modelo para el paciente j recibiendo el tratamiento i es:

$$\begin{aligned} e_{i,j} &\sim Dist(\theta_{e_{ij}}, \sigma_{e_{ij}}) \\ c_{i,j} &\sim Dist(\theta_{c_{ij}}, \sigma_{c_{ij}}) \end{aligned} \quad (1)$$

donde

$$\begin{aligned} \theta_{e_{ij}} &= \mu_{e_i} + \beta_i \cdot (c_{ij} - \theta_{c_{ij}}) + \sum \gamma_e \cdot x_{ij} + \sum \delta_e \cdot I_i \cdot x_{ij} \\ \theta_{c_{ij}} &= \mu_{c_i} + \sum \gamma_c \cdot x_{ij} + \sum \delta_c \cdot I_i \cdot x_{ij} \end{aligned}$$

La distribución que se asume para la efectividad y coste viene descrita por el término *Dist*. La variable I_i se refiere a una variable indicadora de tratamiento que toma valor 1 si el paciente recibe el tratamiento i y 0 en caso contrario.

La modelización anterior asume algunos supuestos discutibles. En primer lugar asume que el efecto de las covariables sobre la efectividad y el coste de todos los tratamientos es igual, con la excepción de los subgrupos definidos previamente. La detección de subgrupos se reduce a la realización de test de hipótesis de nulidad sobre los coeficientes δ que acompañan a las interacciones. Su modelización es apropiada para modelos de distribución normal o gamma pero no tienen aplicación inmediata para otro tipo de distribuciones.

En Moreno *et al.* (2012), se propone un método alternativo de detección de subgrupos a través de métodos de selección de variables Bayesiano. Los métodos de selección de variables para regresión lineal son aplicados utilizando la descomposición de la distribución bivalente del coste y la efectividad. De esta forma se obtienen modelos univariantes del tipo $f(e|\theta_1, \mathbf{x}_{se})$ y $f(c|e, \theta_2, \mathbf{x}_{sc})$, donde $\mathbf{x}_{se} (\subset \mathbf{x})$ hace referencia al conjunto de covariables relevantes para la efectividad y $\mathbf{x}_{sc} (\subset \mathbf{x})$ para el coste.

La principal limitación de esta descomposición es que la igualdad

$$f(c, e|\theta, \mathbf{x}_{sb}) = f(e|\theta_1, \mathbf{x}_{se}) \cdot f(c|e, \theta_2, \mathbf{x}_{sc})$$

no se mantiene necesariamente cuando los subconjuntos \mathbf{x}_{sb} , \mathbf{x}_{se} y \mathbf{x}_{sc} no coinciden.

Posteriormente, Moreno *et al.* (2013) extienden el modelo anterior aplicando un modelo de selección de variables bivalente para la distribución conjunta $f(c, e|\theta, \mathbf{x}_{sb})$. En este trabajo resumimos brevemente dicho proceso.

Tal y como hemos comentado anteriormente, la selección de variables propuesta se realiza desde una perspectiva Bayesiana. La metodología Bayesiana está muy extendida en la literatura coste–efectividad (Al and van Hout, 2000; O’Hagan *et al.*, 2001; Claxton, 1999; Vázquez–Polo *et al.* 2005a,b; Moreno *et al.* 2009, Moreno *et al.* 2012, Moreno *et al.* 2013, among others). Es bastante realista asumir que la información a priori sobre los parámetros del modelo de regresión es escasa, y consecuentemente, adoptamos una perspectiva Bayesiana objetiva para la selección de variables. Para la selección de variables objetiva emplearemos la metodología intrínseca (Berger and Pericchi, 1996; Moreno *et al.*, 1998), adaptada para el caso bivalente. La metodología empleada fue desarrollada recientemente por Torres *et al.* (2011). La construcción de las distribuciones a posteriori y predictivas objetivas son computacionalmente simples para el caso normal–normal y lognormal–normal, que son los que consideraremos en este trabajo.

El resto del documento se organiza de la siguiente forma. En el apartado de Metodología se presenta brevemente el modelo de muestreo Bayesiano bivalente para el coste y la efectividad, el proceso de selección de variables basado en distribuciones a priori intrínsecas, y el ejercicio de simulación en el que compararemos los métodos convencionales de selección de subgrupos con el propuesto por Moreno *et al.* (2013). En el apartado de Resultados se muestran y discuten los resultados del ejercicio de simulación y se añade una aplicación con datos reales del proceso de detección de subgrupos. Finalizará con las conclusiones y futuras ampliaciones del trabajo.

2. Metodología

2.1. Modelo de regresión bivalente para la efectividad y el coste

Denotamos por $f(\mathbf{y}|\theta, \mathbf{x})$ a la distribución bivalente de la efectividad y del coste $\mathbf{y} = (c, e)^t$ para un tratamiento T , donde θ es un vector de parámetros desconocidos, y $\mathbf{x} = (x_1, \dots, x_p)^t$ es un vector de p potenciales covariables. Denotamos por $f(z|R, \theta, \mathbf{x})$ a la distribución del beneficio neto $z = R \cdot e - c$, que se obtiene a partir de la distribución $f(\mathbf{y}|\theta, \mathbf{x})$. Los datos se denotan por $data = (\mathbf{y}_1, \dots, \mathbf{y}_n, \mathbf{X})$, es decir, una muestra de efectividad y coste \mathbf{y} de tamaño n y la matriz de tamaño $n \times p$ de covariables \mathbf{X} .

En la aproximación Bayesiana debemos asumir que θ sigue una distribución a priori $\pi(\theta)$ que actualizaremos con los datos para obtener la distribución a posteriori $\pi(\theta|data)$. Eliminando por integración θ con respecto a su distribución a posteriori $\pi(\theta|data)$ obtenemos

$$f(z|R, data, \mathbf{x}) = \int f(z|R, \theta, \mathbf{x})\pi(\theta|data)d\theta. \quad (2)$$

Esta distribución predictiva del beneficio neto será la medida clave para el problema de decisión bayesiana coste-efectividad.

2.2. Selección de variables objetiva bivalente Bayesiana

En este apartado asumiremos que $f(\mathbf{y}|\theta, \mathbf{x})$ es una densidad normal multivariante, por lo que asumimos que el vector $\mathbf{y} = (c, e)^t$ sigue una distribución normal bivalente $N_2(\mathbf{y}^t|\mathbf{x}^t\mathbf{B}_p, \Sigma_p)$, donde \mathbf{B}_p es una matriz $p \times 2$ de coeficientes del modelo de regresión, y Σ_p una 2×2 matriz de covarianzas.

La verosimilitud de los parámetros (\mathbf{B}_p, Σ_p) para los datos $(\mathbf{Y}, \mathbf{X}_p)$ viene dada por la distribución normal $N_{n \times 2}(\mathbf{Y}|\mathbf{X}_p\mathbf{B}_p, \mathbf{I}_n \otimes \Sigma_p)$.

Por simplicidad restringimos el análisis al caso donde la matriz de covarianzas Σ_p es de la forma $\sigma_p^2 \mathbf{V}$, donde σ_p es un número positivo y \mathbf{V} una matriz simétrica 2×2 . Con este supuesto, el modelo conjunto M_p que contiene las p covariables sería

$$N_{n \times 2}(\mathbf{Y}|\mathbf{X}_p\mathbf{B}_p, \sigma_p^2 (\mathbf{I}_n \otimes \mathbf{V})). \quad (3)$$

Asimismo, la distribución muestral de cualquier submodelo M_j que contenga j regresores vendría dado por la densidad normal $N_{n \times 2}(\mathbf{Y}|\mathbf{X}_j\mathbf{B}_j, \sigma_j^2(\mathbf{I}_n \otimes \mathbf{V}))$, donde \mathbf{X}_j es la matriz de diseño $n \times j$ resultante de eliminar $p - j$ columnas de \mathbf{X}_p , y \mathbf{B}_j en la matriz de coeficientes de dimensiones $j \times 2$. El número de submodelos posibles es 2^p . Si asumimos que la constante siempre está en el modelo, el número de submodelos sería 2^{p-1} .

Debido a la importancia que tienen las covariables sobre la decisión del tratamiento óptimo parece razonable eliminar del modelo completa las covariables que no influyan el coste o la efectividad. Este problema se denomina

selección de variables y consiste en elegir el mejor submodelo M_j de los 2^{p-1} modelos posibles.

Asumiendo la distribución a priori $\pi_j(\mathbf{B}_j, \sigma_j)$ para los parámetros del modelo y una a priori para los modelos $\pi(M_j)$, $j \geq 1$, la probabilidad a posterior para cada modelo sería

$$P(M_j|\mathbf{Y}, \mathbf{X}_j) = \frac{B_{j1}(\mathbf{Y}, \mathbf{X}_j)\pi(M_j)/\pi(M_1)}{1 + \sum_{k=2}^{2^{p-1}} B_{k1}(\mathbf{Y}, \mathbf{X}_k)\pi(M_k)/\pi(M_1)}, \quad (4)$$

donde $B_{k1}(\mathbf{Y}, \mathbf{X}_k)$ es el factor Bayes que compara al modelo M_k frente al modelo que incluye solo la constante M_1 . su expresión es

$$B_{k1}(\mathbf{Y}, \mathbf{X}_k) = \frac{\iint N_{n \times 2}(\mathbf{Y}|\mathbf{X}_k \mathbf{B}_k, \sigma_k^2 (\mathbf{I}_n \otimes \mathbf{V})) \pi_k(\mathbf{B}_k, \sigma_k) d\mathbf{B}_k d\sigma_k}{\iint N_{n \times 2}(\mathbf{Y}|\mathbf{X}_1 \mathbf{B}_1, \sigma_1^2 (\mathbf{I}_n \otimes \mathbf{V})) \pi_1(\mathbf{B}_1, \sigma_1) d\mathbf{B}_1 d\sigma_1}. \quad (5)$$

Aunque la literatura Bayesiana para la selección de variables univariante es amplia (George and McCulloch, 1997; George and Foster, 2000; Clyde and George, 2004; Casella and Moreno, 2006; Liang *et al.*, 2008; Girón and Moreno, 2008), la selección de variables para modelos bivariantes o multivariantes ha tenido poca atención. Afortunadamente, la selección de variables bivalente para a priori intrínsecas ha sido recientemente propuesta por Torres *et al.* (2011).

Utilizando la a priori objetiva uniforme para la probabilidad a priori de los distintos modelos posibles $\{\pi(M_j), j = 1, \dots, 2^{p-1}\}$, la probabilidad a posteriori de cada modelo sería:

$$P(M_j|\mathbf{Y}, \mathbf{X}_j) = \frac{B_{j1}(\mathbf{Y}, \mathbf{X}_j)}{1 + \sum_{k=2}^{2^{p-1}} B_{k1}(\mathbf{Y}, \mathbf{X}_k)}. \quad (6)$$

Sin entrar en detalles técnicos sobre la estructura a priori intrínseca de los coeficientes del modelo (Berger and Perichi, 1996; Moreno *et al.*, 1998; Casella and Moreno (2009), Girón *et al.* (2010), Consonni *et al.* (2011), Leon-Navelo *et al.* (2012), Moreno *et al.* (2012)) se puede demostrar que la expresión del factor bayes para el caso intrínseco sería:

$$B_{k1}(\mathbf{Y}, \mathbf{X}_k) = 2(k+1)^{(k-1)} \int_0^{\pi/2} \frac{\sin(\varphi)^{2(k-1)+1} (n + (k+1) \sin^2 \varphi)^{(n-k)}}{\cos(\varphi)^{-1} [(k+1) \sin^2 \varphi + n\mathcal{B}_{k1}]^{(n-1)}} d\varphi. \quad (7)$$

donde \mathcal{B}_{k1} es el estadístico

$$\mathcal{B}_{k1} = \frac{tr[\mathbf{H}_{\mathbf{X}_k} \mathbf{Y} \mathbf{V}^{-1} \mathbf{Y}^t]}{tr[\mathbf{H}_{\mathbf{X}_1} \mathbf{Y} \mathbf{V}^{-1} \mathbf{Y}^t]},$$

y $\mathbf{H}_{\mathbf{X}} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$.

2.3. Comparación métodos de detección de subgrupos. Ejercicio de simulación.

El objetivo de este trabajo es comparar dos métodos alternativos de detección de subgrupos: el método propuesto por Moreno *et al.* (2013) de selección de variables Bayesiano objetivo y el método convencional de detección de sugrupos a través de test de significación individual sobre los coeficientes de las interacciones tratamiento · subgrupo.

Para realizar la evaluación se desarrolla un ejercicio de simulación con las siguientes características. Tres potenciales covariables (X_1, X_2, X_3) fueron simuladas de distribuciones uniforme entre 0 y 10, $U(0, 10)$. Uno de los objetivos del ejercicio de simulación era detectar la robustez de los métodos empleados a las distintas distribuciones que pueden seguir la efectividad y el coste. Por este motivo se consideró distribución normal para la efectividad y distribución normal, gamma o log-normal para los costes.

$$e_{i,j} \sim N(\theta_{e_{ij}}, 1)$$

$$c_{i,j} \sim \text{NorGammaorLog} - \text{normal}(\theta_{c_{ij}}, 1)$$

La correlación entre la efectividad y el coste se consideró simulando de distribuciones bivariate normal para el caso normal-normal con $\rho = 0,5$ y de una cópula Farlie-Gumbel Morgenstern (FGM) para el caso Normal-Gamma.

Las medias para el caso Normal y Gamma de los dos tratamientos a comparar que fueron asumidas son:

$$\theta_{e_{1i}} = 1 + 0,7 \cdot X_{1i} + 0,2 \cdot X_{2i}$$

$$\theta_{c_{1i}} = 5 + 1 \cdot X_{1i} + 0,3 \cdot X_{2i}$$

para el tratamiento 1 y

$$\theta_{e_{2i}} = 2 + 0,7 \cdot X_{1i} + 0,1 \cdot X_{2i}$$

$$\theta_{c_{2i}} = 8 + 2 \cdot X_{1i} + 0,2 \cdot X_{2i}$$

para el tratamiento 2.

En el caso del modelo Normal-Log-normal se consideró el siguiente modelo para las medias de la transformación logarítmica de los costes que tiene como propiedad que la media coincide con la media de los modelos Normal y Gamma.

$$\theta_{c_{1i}} = 1,74235 + 0,1 \cdot X_{1i} + 0,03 \cdot X_{2i}$$

$$\theta_{c_{2i}} = 1,79444 + 0,2 \cdot X_{1i} + 0,02 \cdot X_{2i}$$

Distintos escenarios para distintos tamaños muestrales fueron considerados. Tamaños muestrales para cada tratamiento que variaban de 50 a 300.

Se realizaron un total de 1000 simulaciones para cada escenario contabilizando el número de veces que el método de detección de subgrupos tomaba la decisión correcta. Como regla de decisión para considerar una elección como correcta se propuso para el modelo de selección de variables objetivo que modelo estimado con mayor probabilidad a posteriori de ser cierto fuese aquel que incluye la constante y las variables X_1 y X_2 , tanto para el tratamiento 1 como para el alternativo.

Por contra, para el modelo propuesto por Nixon and Thompson (2005). En ese caso únicamente la variables X_2 debería ser detectada como variable generadora de subgrupo para la efectividad y las variables X_1 y X_2 como variables que conforman subgrupo en los costes. La estimación también se realizó desde el punto de vista bayesiano para permitir la comparación y se consideró para criterio para considerar un coeficiente como relevante la no inclusión del 0 en la estimación a posteriori del intervalo de credibilidad bayesiano al 95

Las simulaciones fueron realizadas en Mathematika y WinBUGS utilizando el paquete de R denominando R2WinBUGS.

3. Resultados

3.1. Simulación

A continuación se muestran los resultados para los distintos modelos simulados. Las Figura 1-6 muestran la proporción de decisiones correctas para el método de selección de variables objetivo (en azul para el tratamiento 1 y en rojo el tratamiento 2) y el método convencional basado en test de hipótesis sobre los coeficientes interacción (verde para la efectividad y violeta para el coste). En el eje X se muestran las distintas combinaciones de tamaños muestrales consideradas. La Figura 1 muestra el caso en el que los datos han sido simulados de una distribución bivalente Normal-Normal y los métodos de estimación también han supuesto verosimilitudes normales. Se observa como la proporción de decisiones correctas son superiores en el método de selección de variables objetivo que para el método convencional, sea cual sea el tamaño muestral considerado. En el caso del tratamiento 1 la proporción de decisiones correctas del método de selección de variables objetivo supera el 95 % en todos los casos. Sin embargo, con tamaños muestrales pequeños (menores a 100) el método convencional basado el test de nulidad alcanza proporciones de decisiones correctas inferiores al 40 %.

A la derecha de la Figura 1 se muestra el análisis con los mismo datos pero asumiendo distribuciones asimétricas para los costes. En el caso del método de selección de variables se asume una distribución log-normal y en el caso del método convencional se asume una distribución gamma. En este caso se aprecia como el método convencional es muy sensible a una incorrecta especificación del modelo que siguen los datos. Para datos normal

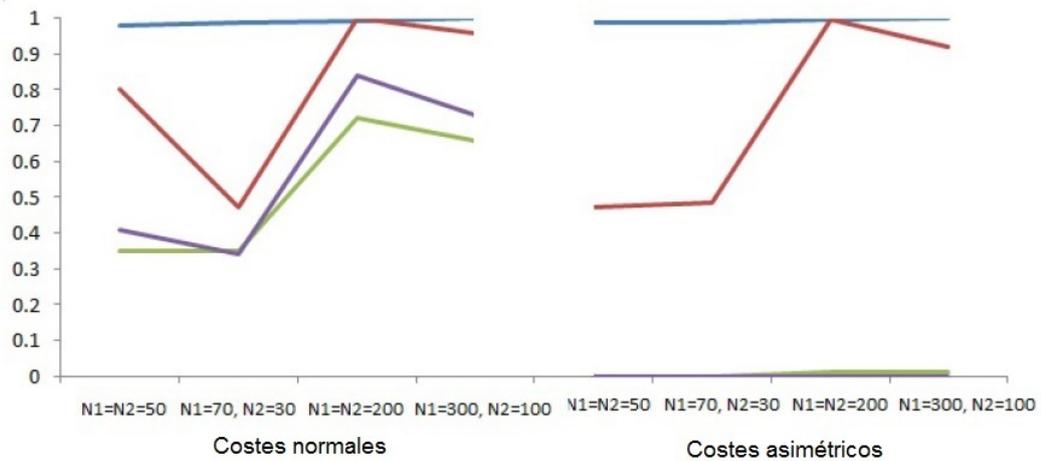


Figura 1: Datos Normales. Proporción de decisiones correctas para el método de selección de variables objetivo (azul para el tratamiento 1 y rojo para el tratamiento 2) y método convencional (verde para la efectividad y violeta para los costes).

y bajo el supuesto de distribución gamma, la proporción de decisiones correctas es muy cercana a 0. Sin embargo, el método de selección de variables objetivo obtienen una proporción elevada de decisiones correctas, al menos, con tamaños muestrales altos.

En la Figura 2 se muestran los resultados del análisis de datos simulados con una distribución normal para la efectividad y gamma para los costes. En este caso nuevamente se observa la sensibilidad del método convencional a una especificación errónea del modelo que siguen los datos. Incluso el modelo que asume distribución gamma para los costes obtienen proporciones de decisiones correctas inferiores al método de selección de variables que asume una distribución log-normal.

Por último, en la Figura 3 se analizan los resultados de los datos de costes simuladas según una distribución log-normal. Nuevamente el método de selección de variables objetivo supera claramente la proporción de éxitos obtenida por los métodos convencionales.

3.2. Ejemplo con datos reales

En este apartado validamos el método de detección de subgrupos propuesto con datos de una prueba clínica real (Hernández *et al.*, 2003) que compara dos tratamientos alternativos para la enfermedad pulmonar obstructiva crónica (EPOC). Se analiza si la hospitalización en casa T_2 para

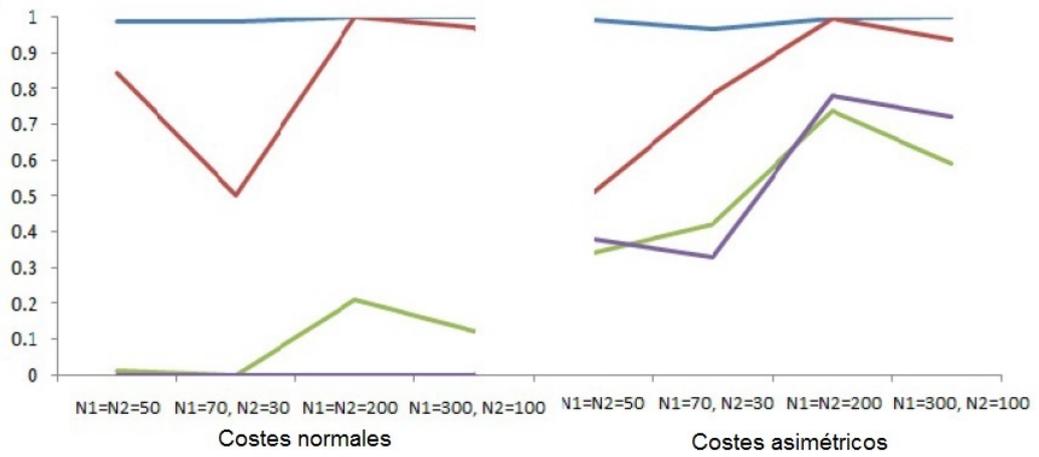


Figura 2: Datos Normal–Gamma. Proporción de decisiones correctas para el método de selección de variables objetivo (azul para el tratamiento 1 y rojo para el tratamiento 2) y método convencional (verde para la efectividad y violeta para los costes).

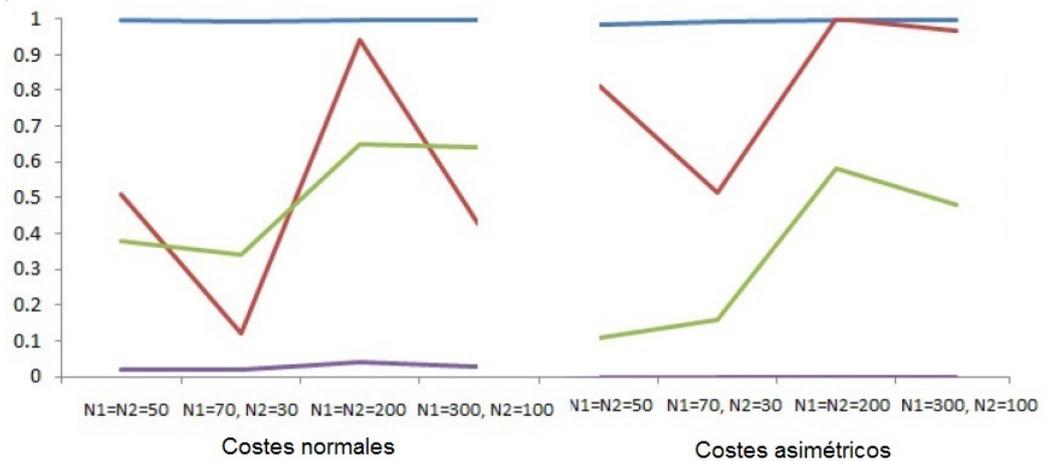


Figura 3: Datos Normal–Log-normal. Proporción de decisiones correctas para el método de selección de variables objetivo (azul para el tratamiento 1 y rojo para el tratamiento 2) y método convencional (verde para la efectividad y violeta para los costes).

pacientes con EPOC es preferida a la hospitalización convencional T_1 . Los pacientes que reciben el tratamiento T_2 tienen un seguimiento continuado de la evolución de la enfermedad con visitas de enfermería al hogar y seguimiento telefónico. Se recabó información de 167 pacientes con EPOC durante 1 año de seguimiento (Noviembre 1999 – Noviembre 2000) en el Hospital Clinic y Hospital de Bellvitge de Barcelona. Los pacientes fueron distribuidos aleatoriamente entre tratamientos, 70 en T_1 y 97 en T_2 .

Las seis variables potenciales de generan subgrupos fueron: edad, sexo, fumar, volumen espiratorio forzado en un segundo (FEV), exacerbaciones que requirieron hospitalización (HOSV) y valor de la escala St. George (SGRQ1) al comienzo del estudio. Como medida de efectividad se consideró la mejora medida a través de la variación en la escala St. George durante 1 año y se incluyeron únicamente costes directos. La Tabla 1 recoge los descriptivos de los datos:

	Tratamiento	
	T_1	T_2
log(Coste)	7,08(1,11)	6,55(0,98)
Efectividad	-1,59(20,15)	7,01(14,07)
N	70	97
Edad	70,44(9,22)	71,28(9,90)
Sexo (% hombres)	97,15 %	97,90 %
Fuma (%)	17,15 %	26,80 %
FEV (%)	39,81 %	43,44 %
HOSV	0,87(1,34)	0,56(0,83)
SGRQ1	50,12(28,01)	54,98(20,2)

Cuadro 1: Media y desviación típica de la efectividad, coste y covariables.

Los resultados de aplicar el modelo de selección bivariante a estos datos nos permiten concluir que las variables influyentes para los costes y la efectividad para el tratamiento 1 son $\{SGRQ1, Age, FEV\}$ y para el tratamiento 2 $\{SGRQ1, FEV\}$.

En presencia de covariables relevantes, el análisis coste-efectividad deberá decidir qué tratamiento es óptimo para las distintas características de los pacientes. La Figura 4 muestra el tratamiento óptimo para distintos niveles de salud inicial, distinta edad y FEV.

De la Figura 4 podemos concluir que para pacientes con mayor edad existe preferencia por la hospitalización en casa. Asimismo, los pacientes con peor estado inicial deben ser tratados a través de la hospitalización convencional.

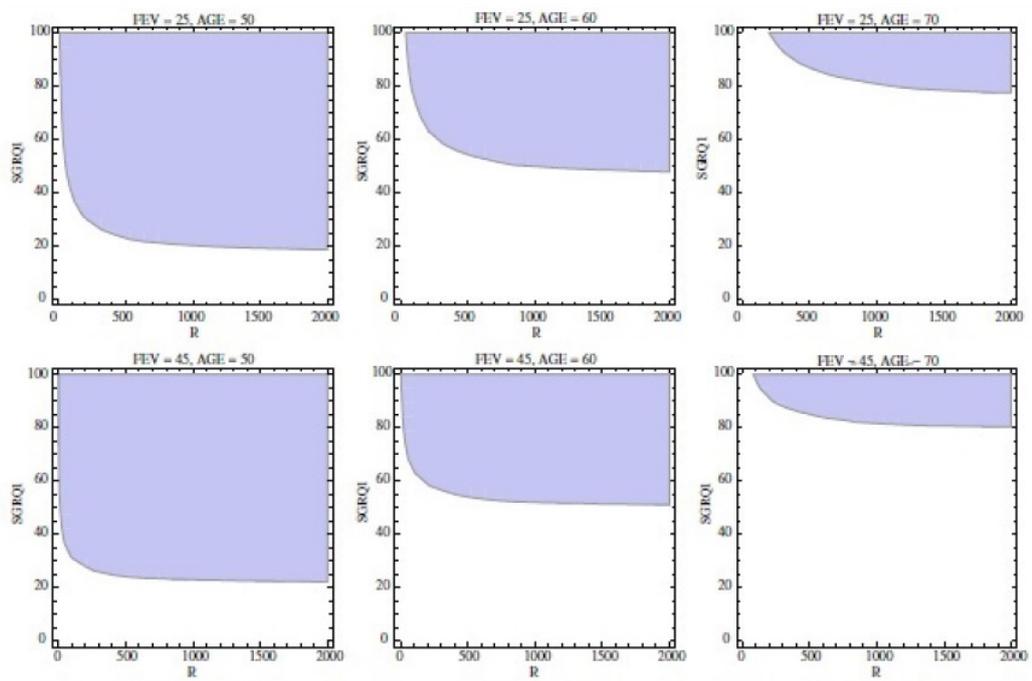


Figura 4: Tratamiento óptimo para distintos valores de las covariables. Color blanco: tratamiento óptimo T_2 y color gris T_1 .

4. Conclusiones

En este trabajo se presenta una manera alternativa de detectar la presencia de subgrupos en el análisis coste–efectividad basada en los métodos de selección de variables Bayesiano. Al realizar el análisis conjuntamente para la efectividad y el coste se ha diseñado un procedimiento para distribuciones bivariantes propuesto muy recientemente en la literatura, aunque aplicado a otro ámbito (Torres *et al.*, 2011).

Se propone un ejercicio de simulación donde se establecen escenarios alternativos tanto en términos de tamaño muestral de los tratamientos como de distribuciones alternativas en la modelización de los costes y la efectividad. La conclusión final ha sido la clara preferencia del método propuesto frente al modelo convencional basado en la inclusión de interacciones entre la variables indicadora de tratamiento y la variable subgrupo.

El método propuesto se muestra robusto tanto a tamaños muestrales pequeños como a distintas distribuciones generadoras de los datos. En ejemplo con datos reales permite destacar las bondades de la metodología propuesta.

Referencias

- Al, M.J. and van Hout, B.A. (2000). A Bayesian approach to economic analysis of clinical trials: the case of Stenting versus Balloon Angioplasty. *Health Economics*, **9**, 599–609.
- Barton, G.R., Briggs A.H. and Fenwick, E.A. (2008). Optimal cost–effectiveness decisions: The role of the cost–effectiveness acceptability curve (CEAC), the cost–effectiveness acceptability frontier (CEAF), and the expected value of perfect information (EVPI). *Value in Health*, **11**, 886–897.
- Berger, J.O. and Pericchi, L.R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91**, 109–122.
- Brailsford, S. and Harper, P. (2008). OR in Health. *European Journal of Operational Research*, **185**, 901–904.
- Casella, G. and Moreno, E. (2006). Objective Bayesian variable selection. *Journal of the American Statistical Association*, **101**, 157–167.
- Casella, G. and Moreno, E. (2009). Assessing robustness of intrinsic test of independence in two–way contingency tables *Journal of the American Statistical Association*, **104**, 1261–1271.
- Casella, G., Girón, F.J., Martínez, M.L. and Moreno, E. (2009). Consistency of Bayesian procedures for variable selection. *Annals of Statistics*, **37**, 731–747.
- Claxton, K. (1999). The irrelevance of inference: a decision making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, **18** (3), 341–364.

- Clyde, M. and George, E.I. (2004). Model uncertainty. *Statistical Science*, **19** (1), 81–94.
- Consonni, G., Moreno, E. and Venturini, S. (2011). Testing Hardy–Weinberg equilibrium: an objective Bayesian analysis. *Statistics in Medicine*, **30**, 62–74.
- Espinoza, M., Manca, A., Claxton, K. and Sculpher, M.J. (2011). The value of heterogeneity for cost–effectiveness subgroup analysis: theoretical framework and application. *Value in Health*, **14** (7), A241.
- George, E. and Foster, D.P. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, **87**, 4, 731–747.
- George, E. I., McCulloch, R.E. (1997). Approaches for variable selection. *Statistica Sinica*, **7**, 339–373.
- Girón, F.J. and Moreno, E. (2008). Comparison of Bayesian objective procedures for variable selection in linear regression. *TEST*, **17**, 472–492.
- Girón, F.J., Moreno, E., Casella, G. and Martínez, M.L. (2010). Consistency of objective Bayes factors for nonnested linear models and increasing model dimension. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas (RACSAM)*, **104**, 161–171.
- Gomes, M., Grieve, R., Nixon, R., Edmunds, W.J. (2012a). Statistical methods for cost–effectiveness analyses that use data from cluster randomized trials: a systematic review and checklist for critical appraisal. *Medical Decision Making*, **32** (1), 209–220.
- Gomes, M., Ng, E.S.W., Grieve, R., Nixon, R., Carpenter, J. and Thompson, S.G. (2012b). Developing appropriate methods for cost–effectiveness analysis of cluster randomized trials. *Medical Decision Making*, **32** (2), 350–361.
- Hernández, C., Casas, A., Escarrabill, J., Alonso, J., Puig–Junoy, J., Farro, E., Vilagut, G., Collvinent, B., Rodríguez–Rosin, R., Roca, J. (2003). Home hospitalization of exacerbated chronic obstructive pulmonary disease patient. *European Respiratory Journal*, **21**, 58–67.
- Leon–Novelo, L., Moreno, E. and Casella, G. (2012). Objective Bayes model selection in probit models. *Statistics in Medicine*, **31**, 353–365.
- Liang, F., Paulo, R., Molina, G., Clyde, M. and Berger, J. (2008). Mixture of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, **103** (481), 410–423.
- Manca, A., Sculpher, M. and Goeree, R. (2010). The analysis of multinational cost–effectiveness data for reimbursement decisions: a critical appraisal of recent methodological developments. *Pharmacoeconomics*, **28** (12), 1079–1096.
- Moreno, E., Bertolino, F. and Racugno, W. (1998). An intrinsic limiting procedure for model selection and hypotheses testing. *Journal of the American Statistical Association*, **93**, 1451–1460.
- Moreno, E., Girón, F.J. and Casella, G. (2010a). Consistency of objective Bayes tests as the model dimension increases. *Annals of Statistics*, **38**, 4,

1937–1952

- Moreno, E., Girón, F.J. and Torres, F. (2003). Intrinsic priors for hypothesis testing in normal regression. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas (RACSAM)*, **97**, 53–61.
- Moreno, E., Girón, F.J., Vázquez–Polo, F.J. and Negrín, M.A. (2009). Complementing the net benefit approach: A new framework for Bayesian cost–effectiveness analysis. *International Journal of Technology Assessment in Health Care*. **25**, 4, 537–545.
- Moreno, E., Girón, F.J., Vázquez–Polo, F.J. and Negrín, M.A. (2010b). Optimal healthcare decisions: comparing medical treatments on a cost–effectiveness basis. *European Journal of Operational Research*, **204**, 1, 180–187.
- Moreno, E., Girón, F.J., Vázquez–Polo, F.J. and Negrín, M.A. (2012). Optimal healthcare decisions: The importance of the covariates in cost–effectiveness analysis. *European Journal of Operational Research*, **218**, 1, 512–522.
- Moreno, E., Girón, L. Martínez, F.J., Vázquez–Polo, F.J. and Negrín, M.A. (2013). Optimal treatments in cost-effectiveness analysis in the presence of covariates: Improving patient subgroup definition. *European Journal of Operational Research*, **226**, 1, 173–182.
- NICE. (2008). Briefing paper for methods review workshop on identifying sub–groups and exploring heterogeneity. National Institute for Health and Clinical Excellence (NICE): London.
<http://www.nice.org.uk/media/4A6/37/SubgroupsBriefingPaper.pdf>
- Nixon, R.M. and Thompson, S.G. (2005). Methods for incorporating covariate adjustment, subgroup analysis and between–centre differences into cost–effectiveness evaluations. *Health Economics*, **14**, 1217–1229.
- O’Hagan, A., Stevens, W. A. and Montmartin, J. (2001). Bayesian cost–effectiveness analysis from clinical trial data. *Statistics in Medicine*, **20**, 733–753.
- Pesch, E. and Woeginger, G. (2012). Operations Research in Health Care. *European Journal of Operational Research*, **219**, 489–490.
- Pinto, J.L., López C., Badía X. Com, A., et al., (2000). Análisis coste–efectividad del tratamiento antirretroviral de gran actividad en pacientes infectados por el VIH asintomáticos. *Medicina Clínica*, **114**, 3, 62–67.
- Pocock, S.J., Assmann, S.E., Enos, L.E. and Kasten, L.E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat. in Medicine*, **21**, 2917–2930.
- Richardson, G. and Manca, A. (2004). Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. *Health Economics*, **13**, 12, 1203–1210.

- Sculpher, M. (2008). Subgroups and heterogeneity in cost–effectiveness analysis. *Pharmacoeconomics*, **26**, 799–806.
- Sculpher, M. and Gafni, A. (2001). Recognising diversity in public preferences: the use of preference sub–groups in cost–effectiveness analysis. *Health Economics*, **10** (4), 317–324.
- Stinnett, A.A. and Mullahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost–effectiveness analysis. *Medical Decision Making*, **18**, S68–S80.
- Torres, F., Moreno, E. and Girón, F.J. (2011). Intrinsic priors for model comparison in multivariate normal regression. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas (RACSAM)*, **105** (2), 273–289.
- Vázquez–Polo, F.J., Negrín, M. and González, B. (2005a). Using covariates to reduce uncertainty in the economic evaluation of clinical trial data. *Health Economics*, **14**, 545–557.
- Vázquez–Polo, F.J., Negrín, M., Badía, X. and Roset, M. (2005b). Bayesian regression models for cost–effectiveness analysis. *The European Journal of Health Economics*, **50**, 45–52.
- Willan, A.R., Briggs, A.H. and Hoch, J.S. (2004). Regression methods for covariate adjustment and subgroup analysis for non–censored cost–effectiveness data. *Health Economics*, **13**, 461–475.