

Metodología de Ajuste de los indicadores del CMBD del SNS. Una técnica genérica basada en Redes Bayesianas

Dr. José Manuel Gutiérrez
Instituto de Física de Cantabria (IFCA)
CSIC - Universidad de Cantabria (UC)
gutierjm@unican.es

Dr. Antonio S. Cofiño
Dpto. de Matemática Aplicada y C.C.
Universidad de Cantabria (UC)
cofinoa@unican.es

CONTENIDOS:

1. Introducción
2. Ajuste por Riesgo (Risk Adjustment)
3. Redes Bayesianas
 - 3.1. Calidad del Modelo: Bondad de Ajuste y Capacidad Predictiva
 - 3.2. Comparación con otras técnicas (regresión logística)
4. Metodología de Ajuste utilizando Redes Bayesianas
5. Conclusiones y Trabajo Futuro

RESUMEN: En este informe se describe la metodología estadística utilizada para ajustar los indicadores del CMBD del SNS acorde a las distintas casuísticas (case-mix) de los pacientes tratados por los distintos proveedores/hospitales (aproximadamente 3,5 millones de pacientes al año tratados en 240 hospitales). Por una parte, estos indicadores están agrupados en ocho grandes familias, incluyendo tanto valores continuos (estancia media, estancia media preoperatoria) como valores binarios en forma de tasas (mortalidad, cesáreas, complicaciones, ambulatorización quirúrgica, reingresos, infección nosocomial). Por otra parte, se dispone de los factores de influencia/riesgo genéricos incluidos en el CMBD para el total de altas: factores relacionados con la propia enfermedad (complejidad, severidad, ROM, CDM, basados en GRDs), con el paciente (edad, sexo) y con los proveedores (tipo de hospital, tipo de ingreso y alta). Se analiza de forma automática la asociación entre los factores y los indicadores (a nivel de paciente y de hospital) considerando el período 2004-2007, aplicando en cada caso el ajuste por aquellos de mayor influencia. Se propone una metodología de ajuste novedosa (las redes probabilísticas) que es común para todos los indicadores y que se considera la más adecuada para el tipo de información disponible (la mayoría de variables involucradas en el problema son discretas). Se muestra que los resultados obtenidos con esta técnica mejoran sensiblemente a los obtenidos con metodologías estándar, basadas en regresión lineal o logística.

En esta fase del proyecto el objetivo ha sido desarrollar una metodología de ajuste común para todos los indicadores, considerando factores de riesgo genéricos para indicadores heterogéneos de distinta naturaleza. En una fase posterior se contempla un análisis más específico de las distintas familias, utilizando factores de comorbilidad específicos para un mejor ajuste en cada una de ellas.

1.Introducción

El registro de altas hospitalarias, también conocido como CMBD (Conjunto Mínimo Básico de Datos al alta), constituye la mayor base de datos administrativa sobre

- | |
|--|
| <ol style="list-style-type: none"> 1. Estancia Media (1) 2. Estancia Media Preoperatoria (1) 3. Tasa de Mortalidad (17) 4. Tasa de Reingresos (1) 5. Tasa de Infección Nosocomial (1) 6. Tasa de Cesáreas (1) 7. Tasa de Complicaciones (15) 8. Tasa de Ambulatorización Quirúrgica (2) 9. Frecuentación en Hospitalización (1) 10. Tasa de Realización (11) |
|--|

Tabla 1. Conjunto de indicadores

pacientes hospitalizados (casi 27 millones de registros, a razón de 3,5 millones/año aproximadamente), siendo la principal fuente de información sobre morbilidad atendida, con información muy valiosa sobre múltiples aspectos de la actividad hospitalaria, incluyendo la calidad y variabilidad de la práctica asistencial.

Esta base de datos recopila la información de 283 hospitales del Sistema Nacional de Salud (SNS).

El Ministerio de Sanidad y Política Social es responsable de la gestión del CMBD estatal, generando con periodicidad anual diversas estadísticas oficiales. Recientemente, se ha desarrollado un modelo de explotación de esta información basado en un conjunto reducido de indicadores de máximo valor explicativo que permita profundizar en el análisis de las características de la atención hospitalaria de los pacientes ingresados en el Sistema Nacional de Salud. Este modelo se basa en 51 indicadores, agrupados en diez familias genéricas mostradas en la Tabla 1 (el número de indicadores en cada familia se muestra entre paréntesis; ver documento de "*descripción del modelo de indicadores*" para más información). Algunos de estos indicadores se refieren a resultados continuos, como número de días, mientras que otros se refieren a resultados discretos, como ocurrencias. Como se verá más adelante, esto obliga a trabajar con técnicas estadísticas distintas en cada caso, complicando el análisis conjunto de los indicadores del CMBD.

Para poder comparar el valor de los indicadores en los distintos hospitales primero es necesario corregir, o ajustar, los resultados brutos teniendo en cuenta las distintas

RELACIONADOS CON LA ENFERMEDAD
<ol style="list-style-type: none"> 1. Complejidad (peso español GRD-AP v18) 2. Severidad (GRD refinados) 3. Riesgo de Mortalidad, ROM 4. Cat. Diagnostica Mayor (GRD-AP v18) 5. Tipo de GRD: médico o quirúrgico
PACIENTE / HOSPITAL
<ol style="list-style-type: none"> 6. Edad 7. Sexo 8. Tipo de ingreso 9. Tipo de alta 10. Tipo de hospital 11. Edad de la madre (sólo para Tasa de Cesáreas)

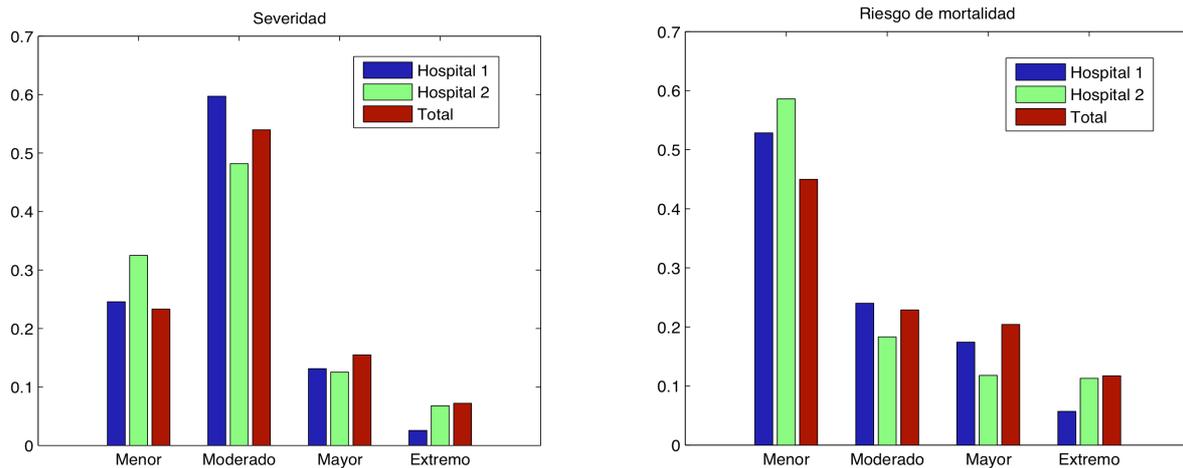
Tabla 2. Factores de riesgo/influencia

casuísticas de los pacientes tratados en los distintos hospitales (case-mix), considerando así la complejidad de los servicios prestados en cada caso.

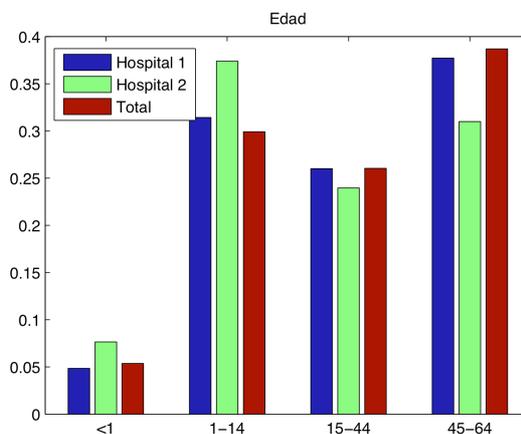
El CMBD incluye un conjunto genérico de factores (denominados factores de riesgo, o influencia, para los indicadores) que dan información de esta casuística y están relacionados tanto con la enfermedad y su diagnóstico (basados en GRDs), como con el paciente y con el funcionamiento hospitalario (Tabla 2). Los factores 5 y 11 son especiales y sólo se utilizan para la infección nosocomial o las tasas de cesáreas, respectivamente.

Estos factores se han utilizado para llevar a cabo una primera corrección genérica de todos los indicadores considerando modelos probabilísticos (redes Bayesianas) que permiten estimar las variaciones que producen los factores de riesgo en un indicador dado; así, se puede ajustar el valor de los distintos hospitales según la casuística particular de los factores de riesgo. Por ejemplo, un incremento de severidad redundará en un aumento de la mortalidad y, por tanto, un hospital que trate un número elevado de altas severas tendrá una mortalidad superior a la de otros hospitales, sin que ello sea debido necesariamente a su funcionamiento, sino sólo a la distinta casuística atendida. Por tanto, el valor bruto del indicador tendrá que ser ajustado (reducido) para tener en cuenta este factor.

Como se verá más adelante, las redes Bayesianas fueron seleccionadas para poder modelar de forma apropiada las complejas distribuciones de los factores de riesgo en los distintos hospitales, así como de sus interacciones conjuntas y de las relaciones con los indicadores. Por ejemplo, la siguiente figura muestra la distribución de *severidad* y *ROM* para dos hospitales distintos del SNS, así como para conjunto total de altas. Se puede observar fácilmente que ambos hospitales tratan pacientes de severidad inferior a la del SNS (la severidad media es similar en ambos casos, 1,93, si se denotan numéricamente en orden creciente los distintos estados) pero con una distribución distinta de casos menores, moderados y extremos.



Por otra parte, si se considera también la distribución de edades, se observa que el Hospital 2 atiende a una población más joven que la media, mientras que el Hospital 1 presenta una distribución de población similar a la del total de altas.



Por tanto, cuando se analiza y compara el resultado particular de un indicador en ambos hospitales es necesario tener en cuenta estas diferencias, de forma que se eliminen los sesgos que son atribuibles a causas/factores conocidos, ajenas a la práctica clínica, y que influyen de forma decisiva en el resultado.

Existen numerosos estudios que analizan los factores de riesgo que pueden ser relevantes para un indicador (ver, por ejemplo, Gruenber et al. 2006 para el caso de estancia media), aunque estos factores pueden ser específicos y no estar disponibles en estudio concreto. Recuérdese que en este trabajo sólo se dispone del conjunto genérico de factores mostrado en la Tabla 2 para todos los indicadores y que, por tanto, el ajuste realizado no pretende ser óptimo para un indicador concreto, sino homogéneo para el conjunto de indicadores considerado.

El ajuste por riesgo (*risk adjustment*) es la denominación genérica para el conjunto de técnicas que tratan de modelizar estadísticamente, a partir de un conjunto de altas, las relaciones existentes entre el indicador y los distintos factores de riesgo, obteniendo modelos que permitan eliminar los sesgos debidos a distribuciones desiguales de factores.

2. Ajuste por Riesgo (Risk Adjustment)

La metodología conocida genéricamente como "ajuste por riesgo" (en ocasiones también denominada ajuste por severidad o ajuste por case-mix) es una parte fundamental de los programas de evaluación de la calidad de la atención sanitaria y es utilizada de manera rutinaria en los distintos proyectos nacionales e internacionales de calidad de proveedores de servicios (ranking, asignación de presupuestos, etc.). Por ejemplo, la AHRQ (www.qualityindicators.ahrq.gov) divulga sus indicadores de calidad acompañados con los resultados de ajustes obtenidos utilizando una serie de factores genéricos de riesgo y su aplicación a los distintos hospitales del sistema de salud. Sin embargo, de esta manera es posible evitar sólo una parte de la variabilidad y es necesario realizar un análisis posterior para su uso en un problema práctico concreto, como la asignación de fondos (ver, por ejemplo, Antioch y Walsh, 2004).

La diversidad de los indicadores utilizados habitualmente (cantidades brutas, medias, tasas, etc.), así como la heterogeneidad de la cantidad y naturaleza de los factores de influencia disponibles (variables categóricas, discretas y continuas), hacen necesario disponer de una batería de métodos estadísticos apropiados para poder aplicar en los distintos casos. De esta manera, en la actualidad existen distintas metodologías que se vienen aplicando a este problema: regresión lineal múltiple, MANOVA, MANCOVA, regresión generalizada --incluyendo la regresión logística--, árboles de regresión y clasificación CART, redes neuronales, etc. (Iezzoni, 2003 y Henderson, 2007 proporcionan una detallada y actualizada introducción a este campo). De entre estas metodologías, la regresión lineal múltiple y la regresión logística son las técnicas más utilizadas para ajustar indicadores continuos (e.g. estancia) y discretos (e.g. mortalidad), respectivamente (ver Hilbe 2009 para una introducción a estas metodologías), y en muchos casos han mostrado una calidad comparable a la de técnicas más complicadas (ver, por ejemplo, Bailit y Garrett, 2003, para un caso de aplicación en tasa de cesáreas). Estas técnicas se suelen emplear en los grandes programas de evaluación de calidad; por ejemplo, estas son las técnicas empleadas por la AHRQ para el ajuste de indicadores (AHRQ, 2007).

El modelo de regresión lineal múltiple se aplica a indicadores continuos, como la estancia, y permite obtener el valor esperado (o predicho) para un paciente dado, Y_k , a partir de los factores de riesgo considerados X_{1k}, \dots, X_{nk} .

$$y_k = a + b_1 x_{1k} + \dots + b_n x_{nk}$$

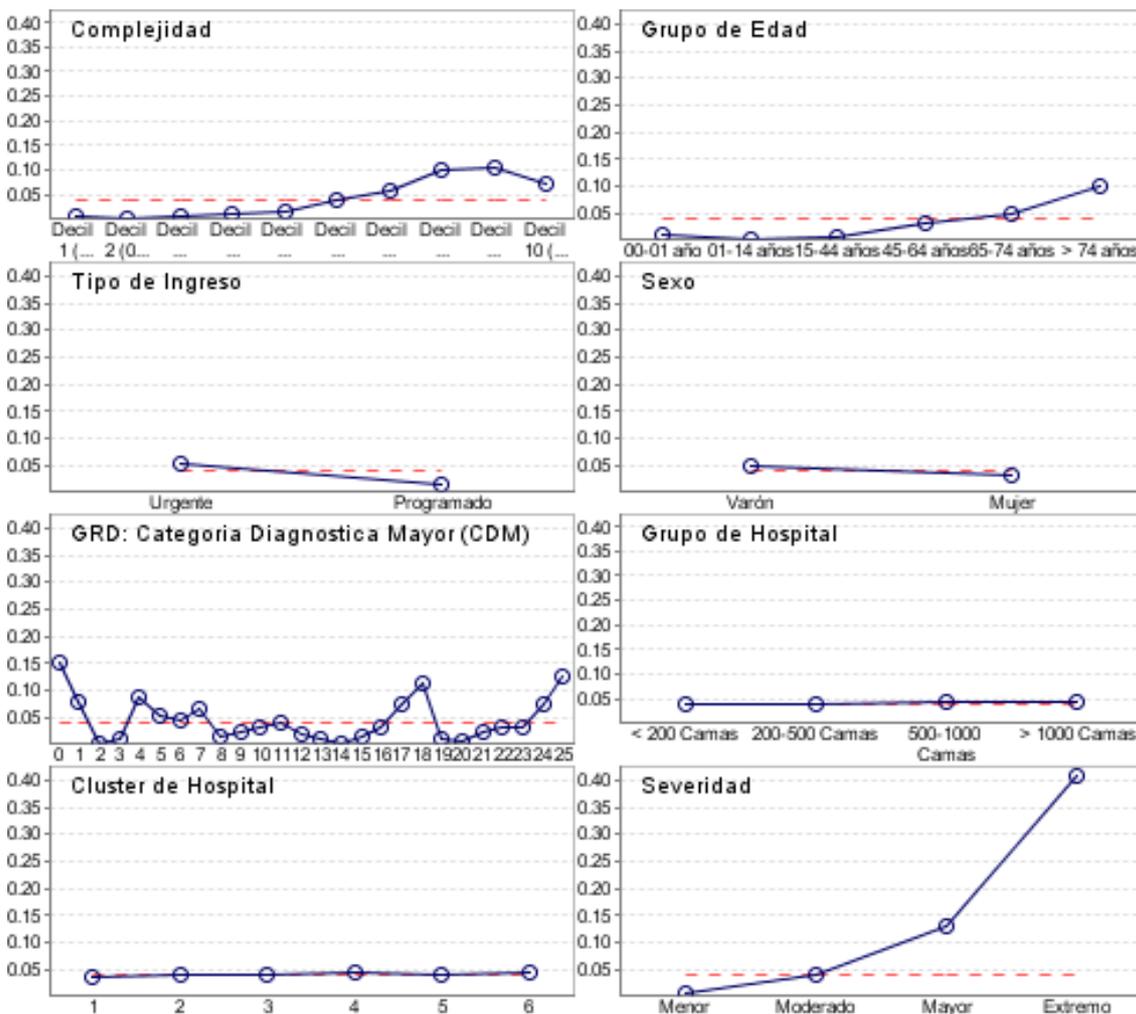
donde los coeficientes a, b_1, \dots, b_n , se ajustan considerando los datos de los pacientes disponibles para optimizar la bondad de ajuste del modelo (ver Hilbe 2009).

La regresión logística es un modelo lineal generalizado que se aplica cuando los resultados son discretos, como la mortalidad; en este caso, el modelo permite obtener la probabilidad de ocurrencia del indicador, entre 0 (no ocurrencia) y 1 (ocurrencia)

$$y_k = 1/1 + \exp(a + b_1 x_{1k} + \dots + b_n x_{nk})$$

Estas técnicas también permiten modelar la variabilidad producida por interacciones entre factores; para ello se pueden introducir co-variables auxiliares que tengan en cuenta el efecto conjunto de pares (e.g. $X_i \cdot X_j$), o tríos, de factores.

La limitación fundamental de estos métodos es que sólo permiten reproducir relaciones lineales (o monótonas para el caso de la regresión logística) entre los factores y el indicador. En algunos casos, esta limitación puede suponer una grave limitación. Por ejemplo, cuando se analiza la variabilidad de la tasa de mortalidad en función de los estados de los distintos factores de riesgo (Tabla 2), se tiene que, en algunos casos, las relaciones son monótonas (crecientes), como el *grupo de edad* o la *severidad*, mientras que en otros casos la relación tiene un máximo, decreciendo ligeramente en los casos más extremos, como la *complejidad*.



En este caso, resultaría razonable realizar un ajuste de riesgo aplicando un modelo de regresión logística considerando como factores de riesgo el *grupo de edad*, la *severidad* y la *complejidad*. Sin embargo, el modelo sería aproximado y proporcionaría una probabilidad creciente de mortalidad al incrementar la complejidad.

Por otra parte, la regresión lineal y logística son más apropiadas cuando los factores disponibles son variables continuas. Sin embargo, los factores mostrados en la Tabla 2 son variables discretas con la excepción de la complejidad y la edad; esta última variable se suele discretizar de forma natural en distintos estratos de edades. La complejidad presenta una compleja distribución y, por ello, en este trabajo también se considera de forma discretizada (ver documentación técnica del análisis descriptivo de factores de influencia). Por tanto, todos los factores considerados en este trabajo han sido factores discretos.

Los modelos gráficos probabilísticos (en particular las redes Bayesianas) constituyen una metodología alternativa a la regresión para trabajar con variables discretas, permitiendo modelizar una función de ajuste genérica (lineal o no lineal) y considerando los efectos relevantes entre conjuntos de factores para la variabilidad del indicador (ver Castillo et al. 2004 para una introducción a las redes probabilísticas). Estos modelos fueron desarrollados en la década de los ochenta y se utilizan profusamente en ámbitos tan variados como el marketing y la genética, habiendo comenzado a utilizarse recientemente en algunas aplicaciones médicas (ver, por ejemplo Acid et al. 2004).

Así como la regresión es una técnica estadística clásica, las redes probabilísticas pertenecen a la disciplina denominada minería de datos, que engloba técnicas estadísticas avanzadas que pueden aplicarse de forma eficiente para modelizar y analizar grandes conjuntos de datos (como el problema que nos ocupa). Estas técnicas suelen ser más complejas y su característica principal es que permiten obtener de forma automática información que está escondida en los datos y que forma la base para construir un modelo explicativo subyacente. En el caso de las redes Bayesianas, el conocimiento oculto que se obtiene son las relaciones de dependencia e independencia relevantes entre conjuntos de variables (por ejemplo, las interacciones entre factores), permitiendo conocer cómo se relacionan globalmente los factores de riesgo entre ellos y con el indicador, así como construir un modelo lo más sencillo posible que incluya únicamente las relaciones relevantes.

Cuando las variables/factores son discretas, las redes Bayesianas son la solución más eficiente para construir un modelo probabilístico conjunto de todas las variables, permitiendo estimar la probabilidad del indicador a partir de los factores de riesgo conocidos: $P(y_k, x_{1k}, x_{nk})$. En este caso, la red Bayesiana ofrece una alternativa eficiente para ajustar el indicador según las distintas combinaciones de factores que definen los posibles perfiles y sus probabilidades.

3.Redes Bayesianas

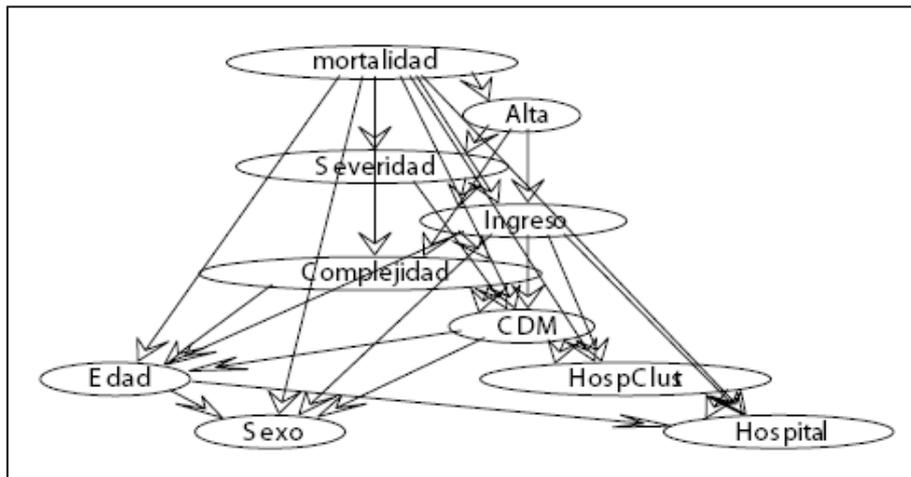
Cuando todos los factores de riesgo son variables discretas, entonces la regresión simplemente ajusta el indicador considerando el valor promedio poblacional en cada uno de los estados de los factores de riesgo. En este caso se agravan las limitaciones del método al no poder considerar efectos conjuntos de los factores de riesgo y no poder modelizar relaciones no lineales. Para poder definir un modelo de ajuste más general es necesario trabajar con la probabilidad conjunta de todas las variables:

$$P(y, x_1, x_n) = P(\text{indicador}, \text{cdm}, \text{complejidad}, \text{sexo}, \text{edad}, \text{ingreso}, \text{hospital}, \text{alta}, \text{severidad}).$$

Sin embargo, la especificación completa de esta probabilidad conjunta requeriría más de 50 millones de parámetros (el producto de todos los posibles estados de las variables, para tener en cuenta todas las interacciones). Por tanto, se hace necesario simplificar esta función eliminando aquellas dependencias que no sean relevantes en el conjunto de datos a estudiar y que permitan definir la probabilidad conjunta con un número menor de parámetros. Para ello, se han considerado los modelos gráficos probabilísticos (en concreto las redes Bayesianas) como una metodología sólida y extendida para abordar este problema (Castillo et al. 2004).

Estos modelos utilizan un grafo (dirigido acíclico) para representar de forma cualitativa las dependencias relevantes entre variables (cada enlace en el grafo indica una dependencia directa entre variables y los distintos caminos indican dependencias indirectas, según un criterio conocido como criterio de *d-separación*). El grafo resultante permite definir la probabilidad conjunta de forma simplificada a partir del producto de las probabilidades condicionadas de cada variable dados sus padres en el grafo (variables que apuntan a la variable en cuestión). Esta factorización permite expresar la probabilidad conjunta de forma compacta, requiriendo un número reducido de parámetros y conservando todas las dependencias representadas por el grafo.

Para definir una red Bayesiana para cada indicador se ha procedido en dos etapas. Primero, dado que el indicador es la variable objetivo, se parte de una estructura de clasificador Bayes ingenuo (el indicador es padre de todos los factores); de esa forma se garantiza que las dependencias directas entre el indicador y cada uno de los factores de influencia estarán caracterizadas de forma exacta en el modelo resultante. A continuación, aplicando una técnica de aprendizaje automático, se infieren las dependencias relevantes entre los factores a partir del conjunto de altas disponible. Para ello se han utilizado y comparado distintos algoritmos de aprendizaje automático, basados en una búsqueda iterativa del grafo óptimo. Tras las pruebas realizadas se ha elegido un algoritmo voraz denominado “algoritmo B”, que va añadiendo sucesivamente los enlaces que mejor explican los datos (altas), penalizando modelos demasiado complejos que se pueden sobreajustar a los datos disponibles (una descripción completa de los métodos de aprendizaje puede consultarse en Castillo et al. 1997; <http://personales.unican.es/gutierjm/BookCGH.html>). Por ejemplo, el grafo obtenido para el indicador *tasa de mortalidad* se muestra en la siguiente figura.



El modelo probabilístico resultante permite calcular la probabilidad a posteriori del indicador (e.g. probabilidad de *exitus*) a partir de los estados concretos de los factores

de riesgo (e.g. *severidad extrema*). Por tanto, se puede trabajar a nivel de paciente y estimar la ocurrencia del indicador a partir de los factores de riesgo correspondientes, o a nivel de hospital, estimando la tasa de ocurrencia del indicador a partir de las distribuciones de los factores de riesgo en el hospital (e.g. 70% *menores*, 15% *moderados*, 10% *mayores* y 5% *extremos*). El valor estimado es el que correspondería a la situación del paciente o del hospital, inferido por el modelo estadístico a partir de los factores de riesgo en base al total de altas disponible.

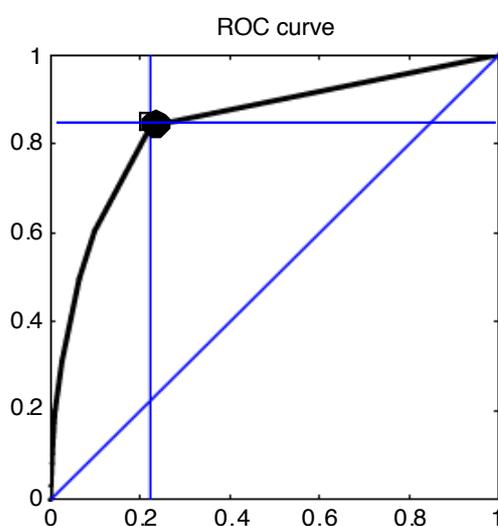
3.1. CALIDAD DEL MODELO: BONDAD DE AJUSTE Y CAPACIDAD PREDICTIVA

Para evaluar la calidad de los modelos de ajuste resultantes, de forma que se pueda comprobar el grado de ajuste de los indicadores y comparar otros métodos, se han utilizado medidas estándar según el indicador sea una media o una tasa (ver Iezzoni, 2003, para más detalles).

Cuando el indicador es una variable continua (indicadores 1 y 2 en la Tabla 1) las medidas de calidad más habituales se basan en medidas de bondad de ajuste, como el porcentaje de varianza explicada, dado por el estadístico R^2 , que indica la fracción de variabilidad del indicador en la población que puede ser eliminada por causas conocidas por los factores de riesgo utilizados en el ajuste. De esta manera, cuanto mayor sea el porcentaje de varianza explicada, mejor será el ajuste del indicador.

Cuando el indicador es una tasa (indicadores 3-8 en la Tabla 1), las medidas de evaluación son más complejas porque el modelo proporciona la probabilidad de la ocurrencia del indicador para cada paciente, a partir de los valores de los factores de riesgo. Por tanto, es necesario utilizar medidas de evaluación para predicciones probabilísticas que se basan en el concepto de *capacidad predictiva* en lugar de en *bondad de ajuste*. Las medidas de capacidad predictiva más comunes se basan en las tablas de contingencia obtenidas para distintos umbrales de la probabilidad, caracterizadas por los valores de *sensibilidad* (probabilidad de detectar la ocurrencia si se ha producido), y *especificidad* (probabilidad de detectar una no ocurrencia) del método. La curva ROC se construye en base a estos valores y el área encerrada por la misma (AUC, *area under ROC curve*) caracteriza la capacidad predictiva del método.

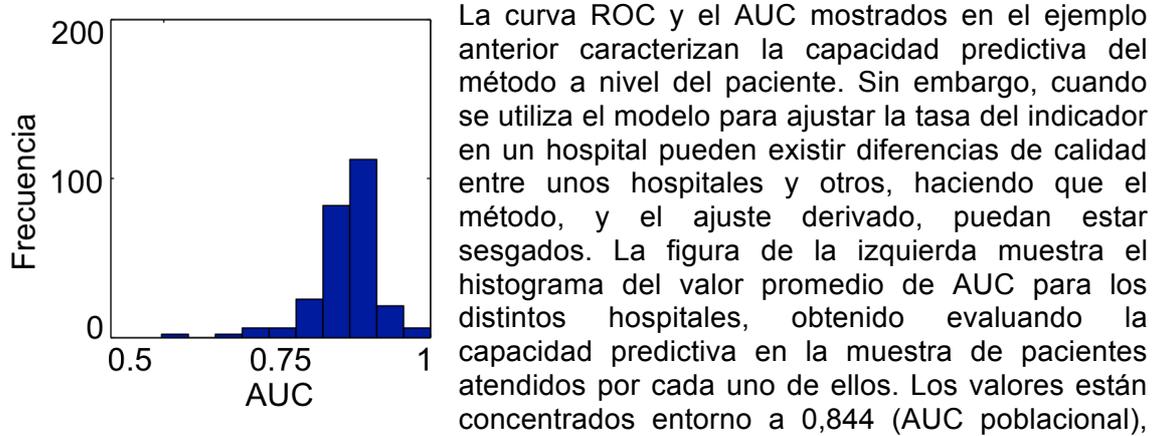
Una curva ROC enfrenta en un sistema de ejes la sensibilidad (en el eje y), al complementario de la especificidad (en el eje x). El procedimiento para construir la



curva consiste en utilizar un conjunto creciente de umbrales de probabilidad 0,1, 0,2, ..., 1 (se considera la ocurrencia cuando la probabilidad prevista es superior al umbral), y determinar los valores de sensibilidad y especificidad en cada caso. El valor de AUC oscila entre 0 y 1, con valores mayores para métodos con mayor capacidad predictiva (un método perfecto tiene un AUC de 1); en la práctica se suele considerar el valor 0,6 como un indicador mínimo de calidad.

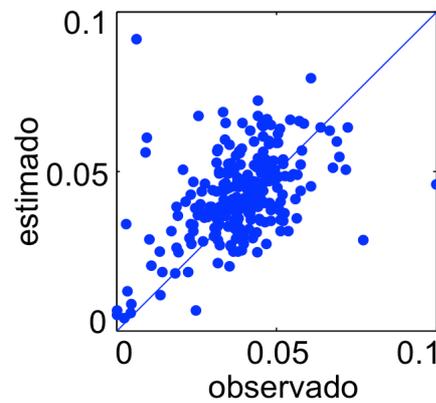
Por ejemplo, la figura izquierda muestra la curva ROC obtenida al aplicar una red probabilística a la tasa de mortalidad, considerando los factores de riesgo *grupo de edad*, *severidad* y *complejidad*, y la población de pacientes del año 2007. El valor del área bajo esta curva es de $AUC=0,844$, lo que

caracteriza la capacidad predictiva del modelo. Por ejemplo, el punto mostrado en la figura indica que la red Bayesiana es capaz de predecir acertadamente, a partir de los factores de riesgo, más de un 80% de los casos (exitus de un paciente), cometiendo únicamente un 20% de falsas predicciones.

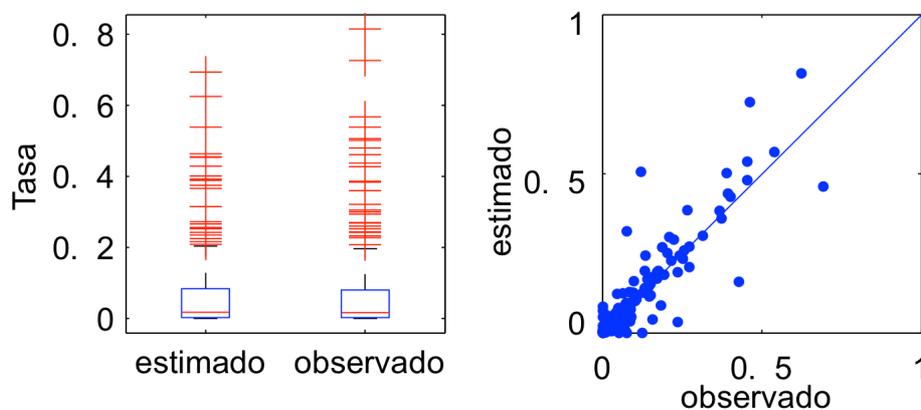


siendo superiores a 0,6, indicando una buena calidad predictiva en todos los casos.

La figura de la derecha muestra las tasas de mortalidad observadas frente a las estimadas en los distintos hospitales del SNS. Esta figura muestra una clara relación entre ambas magnitudes con una ligera tendencia al incremento/decremento del valor estimado para hospitales con valores pequeños/grandes de la tasa observada. Estos resultados forman la base de la metodología de ajuste que se describe más adelante.



Por otra parte, también es deseable que la capacidad predictiva del modelo sea similar para todas las asociaciones de los valores de los factores de riesgo, o grupos de riesgo. En caso contrario, el valor ajustado de un hospital podría estar sesgado por una mala capacidad predictiva para su situación particular. Las figuras inferiores muestran las tasas estimadas y observadas en los 240 grupos de riesgo distintos para *severidad*, *complejidad* y *edad* (los factores utilizados para ajustar el valor de este indicador, como se describe más adelante).



Los indicadores analizados en esta primera fase del trabajo son los indicadores principales numerados del 1 al 8 en la Tabla 1; los indicadores 9 y 10 requieren un

análisis especial pues el denominador de la tasa es variable (poblaciones de comunidades autónomas, etc.). En una fase posterior se prevé abordar el estudio de los subindicadores, considerando sus distintas particularidades.

La Tabla 3, mostrada a continuación, resume la bondad de ajuste, o la capacidad predictiva, según corresponda, de los modelos de red Bayesiana obtenidos utilizando la metodología propuesta para el período 2004-2007; esta tabla indica que, en la mayoría de los casos, no existe una variabilidad significativa de la calidad de unos años a otros.

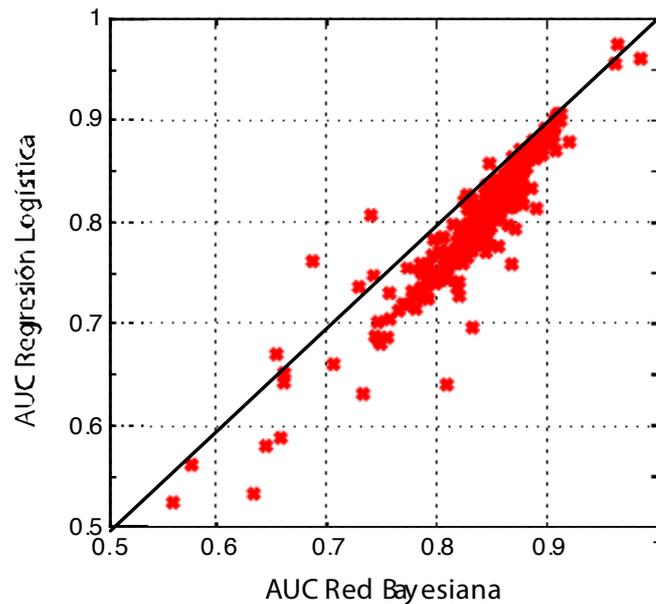
Indicador (medias), R²	2004	2005	2006	2007
Estancia Media	39,2	37,2	42,5	39,8
Estancia Media Preoperatoria	28,7	30,4	35,2	35,6
Indicador (tasas), AUC	2004	2005	2006	2007
Tasa de Mortalidad	0,847	0,849	0,845	0,843
Tasa de Reingresos (a 30 días)	0,652	0,657	0,660	0,667
Tasa de Infección Nosocomial	0,665	0,666	0,658	0,664
Tasa de Cesáreas	0,547	0,545	0,546	0,549
Tasa de Complicaciones	0,684	0,700	0,721	0,720
Tasa de Ambulatorización Quirúrgica	0,844	0,837	0,835	0,830
Tabla 3. Medidas de Calidad de los Modelos				

3.2. COMPARACIÓN CON OTRAS TÉCNICAS (REGRESIÓN LOGÍSTICA)

Con objeto de poder comparar la capacidad predictiva de las redes Bayesianas con otros modelos estándar, se ha realizado un estudio considerando un indicador típico en forma de tasa (la tasa de mortalidad), para el que se han construido un modelo de ajuste basado en redes Bayesianas, y otro basado en regresión logística, considerando en ambos casos los mismos factores de ajuste y la misma muestra de individuos (3.686.655 altas del año 2007).

Los resultados han sido agregados por hospitales para evaluar la variabilidad de la capacidad predictiva del método de ajuste en el conjunto de hospitales. En la figura inferior se resumen los resultados hospital por hospital (puntos rojos), mostrando los valores de capacidad predictiva (AUC) obtenidos utilizando una red Bayesiana, frente a los obtenidos con una regresión logística. En ambos casos, la variabilidad de AUC oscila en su mayoría entre 0,75 y 0,9 (según los hospitales), con un valor promedio de 0,843 y 0,799 respectivamente.

Por tanto, aunque existe una cierta variabilidad en los resultados, ambos métodos tienen una capacidad predictiva elevada en todos los hospitales. Esta figura también muestra que ambos métodos obtienen una corrección similar (existe una clara relación lineal entre ambas; es decir, valores altos/bajos de un modelo se corresponden con valor altos/bajos del otro). Sin embargo, la figura muestra la superioridad de la red Bayesiana frente a la regresión logística, ya que casi todos los hospitales están a la derecha de la diagonal. No obstante, el modelo logístico podría mejorarse considerando efectos conjuntos como factores adicionales del modelo. Sin embargo, siempre existiría la limitación impuesta por la imposibilidad del modelo para modelizar relaciones no lineales (por ejemplo, relaciones no monótonas).



La comparación anterior se muestra para poner de manifiesto que la nueva metodología de ajuste propuesta en este trabajo sigue unas pautas similares a las de otros métodos estándar que se aplican a este problema de manera habitual.

4. Metodología de Ajuste utilizando Redes Bayesianas

Las redes Bayesianas, descritas en la Sección 3, permiten calcular de manera eficiente las probabilidades $P(\text{indicador}=x \mid \text{factores}=\text{valores})$ para los distintos estados x del indicador dados los estados de los factores presentes en el paciente (por ejemplo, *severidad = extrema*), o las distribuciones de factores dados en el hospital (por ejemplo, $P(\text{severidad} = \text{extrema})=0,1$). Si el indicador es una tasa o proporción (indicadores 3 al 8 de la Tabla 1), dado por una variable binaria (por ejemplo, *exitus=si* o *no*), entonces la probabilidad esperada del paciente viene dada directamente por:

$$\text{Valor esperado} = P(\text{indicador}=\text{si} \mid \text{factores}=\text{valores}).$$

Si el indicador es una variable discreta (como los indicadores 1 y 2 de la Tabla 1: estancia media y estancia media preoperatoria, que fueron discretizados según se describe en el informe de "análisis descriptivo de los indicadores"), entonces el valor esperado puede obtenerse de la siguiente forma:

$$\text{Valor esperado} = \sum_x x * P(\text{indicador}=x \mid \text{factores}=\text{valores}).$$

Por tanto, para obtener el valor esperado de un indicador en un hospital concreto, es necesario calcular la distribución a posteriori de Y a partir de la casuística particular de los factores; en términos Bayesianos, esta casuística viene dada por la distribución a priori de los factores en el hospital: $P_h(X=x) := P(X=x \mid H=h)$, que se combina con la probabilidad del indicador dados los factores de influencia, $P(Y=y \mid X=x)$, para obtener la probabilidad a posteriori del indicador condicionada a la casuística del hospital (reflejada en los factores de influencia):

$$P_h(Y=y) := P(Y=y \mid H=h) = \sum_x P(Y=y \mid X=x) * P_h(X=x). \quad (5)$$

$P_h(Y=y)$ es la distribución de probabilidad del indicador que correspondería a la casuística de los factores de influencia concretos de un hospital $H=h$. Por tanto, la diferencia del valor $E_h(Y) = \sum_y y \cdot P_h(Y=y)$ respecto del valor medio poblacional $E(Y)$ indica la variación del valor del indicador del hospital H que puede explicarse en base a su casuística particular de los factores relevantes considerados. Por tanto, para corregir el valor del indicador observado en el hospital habrá de calcularse:

$$YA_h = Y_h + (E(Y) - E_h(Y)) = E(Y) + (Y_h - E_h(Y)), \quad (6)$$

siendo Y_h el valor original del indicador del hospital, y siendo YA_h el valor ajustado eliminando la influencia de los factores relevantes. Por tanto, las ecuaciones anteriores proporcionan la metodología de ajuste o corrección, y la red Bayesiana el mecanismo para llevar a cabo los cálculos involucrados. De esta forma, si las características de un hospital (en base a los factores relevantes) le perjudican elevando el valor de su indicador respecto del valor medio poblacional, entonces el factor de corrección reduce este valor eliminando el perjuicio. De forma análoga, cuando las características de un hospital favorecen al indicador, el factor de corrección actúa incrementándolo en la medida que puede explicarse a partir de sus características.

Para llevar a cabo el ajuste de cada uno de los indicadores de la Tabla 1, se entrenó una red Bayesiana con el conjunto total de altas, considerando como variables el indicador y los posibles factores de influencia (ver Sección 3), excluyendo aquellos factores que pudieran considerarse efectos, y no causas, del indicador. La Tabla 4 muestra (en las celdas sombreadas) los factores utilizados en cada caso.

A continuación, para proceder al ajuste de los indicadores utilizando (5) y (6) evitando el sobreajuste se eligió la combinación de un reducido número de factores de ajuste X que maximizasen la varianza explicada (o el AUC) del modelo, incluyendo el mínimo número de parámetros; para penalizar el número de parámetros se utilizó un término de penalización basado en la medida de información de Akaike (ver Castillo y otros 2007). La Tabla 4 muestra, en color rojo, los factores de ajuste que se consideraron en cada caso; en la mayoría de los indicadores el óptimo se alcanza con tres factores.

	Severidad	ROM	Sexo	CDM	Edad	Complejidad	Tipo Ingreso	Hospital	Tipo Alta	Tipo GRD	Edad madre
Estancia											
Est. preop.											
Mortalidad											
Reingresos											
Infección											
Cesáreas											
Complica.											
Ambulat.											

Tabla 4. Indicadores y factores de riesgo considerados en el ajuste

Obsérvese que aunque sólo se considera un conjunto reducido de factores de ajuste, la red Bayesiana incluye la información de las relaciones con los restantes factores y, por tanto, el ajuste final tiene información de todos los factores, directa o indirectamente.

5. Conclusiones y Trabajo Futuro

El objetivo de este trabajo era desarrollar y validar una metodología de ajuste por factores de riesgo que pudiese aplicarse de forma genérica, considerando un conjunto común de factores de riesgo, a todos los indicadores definidos por el Instituto de Información Sanitaria (IIS) a partir del CMBD (51 indicadores, agrupados en diez familias). La conclusión principal de este trabajo es que las redes probabilísticas ofrecen una metodología de ajuste apropiada para este problema, ya que la mayoría de factores disponibles son variables discretas (a excepción de complejidad). Esta técnica permite ajustar indicadores en forma de medias y de tasas, ofreciendo unos resultados superiores a los de metodologías estándar utilizadas en otros proyectos similares (regresión lineal múltiple y regresión logística). La siguiente tabla resume los resultados obtenidos para los ocho indicadores genéricos de las familias, mostrados en la Tabla 1.

En algunos de los casos, los modelos de ajuste obtenidos en este trabajo tienen una calidad similar a la de otros estudios particulares más exhaustivos para un indicador, que han utilizado factores de riesgo específicos para el indicador tratado. Por ejemplo, el modelo obtenido en este trabajo para la estancia media permite explicar una varianza de entre 25 y 30% (según el año), que son valores similares a los encontrados en la literatura en distintos estudios. También se pueden encontrar en la literatura índices de capacidad predictiva AUC para la mortalidad del orden de 0,85 (ver, por ejemplo, Stukenborg et al. 2001, que utiliza distintos factores de comorbilidad y administrativos). Por otra parte, tal como se muestra en la Tabla 3, aplicando esta metodología al total de hospitales del SNS, considerando los factores de riesgo disponibles, se obtuvo un modelo de ajuste para la tasa de cesáreas con una calidad predictiva dada por un AUC de 0,55. En este caso, en la literatura existen trabajos con mejores índices de calidad, que trabajan en un entorno más favorable y con más información. Por ejemplo, Peaceman et al. (2002) obtienen un AUC de 0,86 para la tasa de cesáreas considerando un único hospital universitario, aplicando un modelo de regresión logística con 15 factores de riesgo, alguno de ellos específico para este indicador (peso al nacer, tiempo de gestación, etc.). La pérdida de calidad en los modelos de ajuste cuando no se dispone de factores de riesgo específicos ha sido descrita en la literatura para otros indicadores particulares como la mortalidad por infarto (Krauss y Cruz, 2000), etc.

Esto muestra que, una vez terminada la primera fase del proyecto, conducente a obtener una metodología común y genérica para el ajuste de los indicadores, es posible analizar de manera más exhaustiva algunas familias de indicadores que admiten una mejora.

Referencias

- E. Castillo, A.S. Hadi, and J.M. Gutiérrez. *Expert Systems and Probabilistic Network Models*. Springer. 1997
- S. Acid, L.M. De Campos, J.M. Fernández-Luna et al. (2004) A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial Intelligence in Medicine*, 30, 215–232.

J.M. Gutiérrez et al. (2007) Measuring hospital outcomes: Statistical model to determine variables of influence for adjustment in Spanish public hospitals. *23rd PCS International Working Conference - 7/10 November 2007, Venice, Italy*

M.A. Gogorcena et al. (2007). Model of indicators based on Spanish NHS Hospital Discharge Minimum Basic Data Set (MBDS). *23rd PCS International Working Conference - 7/10 November 2007, Venice, Italy*

AHRQ (2007) US Department of Health and Human Services, Agency for Healthcare Research and Quality. AHRQ Quality Indicators: Inpatient Quality Indicators: Technical Specifications: Covariates. March 2007. Available at: http://www.qualityindicators.ahrq.gov/downloads/iqi/iqi_covariates_v31.pdf

W.G. Henderson S.F. Khuri. *Risk Adjustment*. In Risk Adjustment in Clinical Research Methods for Surgeons, Edited by D. F. Penson and J. T. Wei, 105-122, Humana Press Inc. 2007.

L.I. Iezzoni. *Risk Adjustment for Measuring Health Care Outcomes*, 3rd Ed. Health Administration Press, 2003.

D.A. Gruenberg et al. (2006) Factors influencing length of stay in the intensive care unit. *American Journal of Critical Care*, 15, 502-509.

J.M. Hilbe. *Logistic Regression Models*. Chapman & Hall/CRC Press. 2009.

K. Antioch and M. Walsh (2004) The risk-adjusted vision beyond casemix (DRG) funding in Australia: International lessons in high complexity and capitation. *The European Journal of Health Economics*, 5, 95-109

L. Krauss and F.O. Cruz (2000) Validity of the risk adjustment approach to assess effectiveness and improving outcomes in Rio De Janeiro. *Annu Meet Int Soc Technol Assess Health Care Int Soc Technol Assess Health Care*, 16: 324.

G. Stukenborg et al. (2001) A New Risk Adjustment Method for Use with Hospital Administrative Data. *Abstr Acad Health Serv Res Health Policy Meet*. 18, 186.

M.S. Hendryx et a. (2001). An Examination of Methods for Risk-Adjustment of Rehospitalization Rates. *Mental Health Services Research*, 3, 15-24.

J. Bailit and J. Garrett (2003) Comparison of Risk-Adjustment Methodologies for Cesarean Delivery Rates. *Obstetrics & Gynecology*, 102, 45-51.