

## 5.5 Métodos para medir el desempeño de los centros asistenciales



La evaluación de centros y servicios sanitarios exige instrumentos que integren dimensiones de actividad y resultados y lo relacionen con los recursos; en este tema se exponen las principales tipologías de análisis de desempeño.

**Autores: Beatriz González López-Valcárcel(1)  
José Ramón Repullo Labrador(2)**

(1) *Catedrática de Métodos Cuantitativos en Economía y Gestión  
Universidad de Las Palmas de GC*

(2) *Jefe de Departamento de Planificación y Economía de la Salud. ENS. Madrid*

Se recomienda imprimir 2 páginas por hoja

### Citación recomendada:

González López-Valcarcel B. Repullo Labrador J R. Métodos para medir el desempeño de los centros asistenciales[Internet]. Madrid: Escuela Nacional de Sanidad; 2012. Actualizado feb 2022 [consultado día mes año]. Tema 5.5. Disponible en: direccion url del pdf.



TEXTOS DE ADMINISTRACION SANITARIA Y GESTIÓN CLINICA  
by UNED Y ESCUELA NACIONAL DE SANIDAD  
is licensed under a Creative Commons  
Reconocimiento- No comercial-Sin obra Derivada  
3.0 Unported License.



### Resumen:

Los métodos para obtener un ranking de centros por desempeño difieren según los objetivos, los datos y el modelo subyacente (grado de incertidumbre y forma de incorporar, en su caso, juicios de valor para ponderar diferentes objetivos).

El ranking de criterio único incluyen los modelos econométricos de frontera estocástica, que miden comparativamente la eficiencia de las unidades productivas en el marco teórico de la teoría económica de la producción; y los modelos de regresión logística

para estimar la calidad, que se ajustan por factores de confusión y riesgos.

### *Introducción. Ranking y Benchmarking*

#### *1- ¿Qué se entiende por desempeño?*

*1.1-Inputs, outputs y funciones de producción*

*1.2- Algunas experiencias a título ilustrativo*

#### *2 - Tipología de los métodos*

#### *3- Los ranking unidimensionales (basados en una medida única de "éxito")*

*3.1- Fases del proceso de construcción del ranking unidimensional de centros*

*3.2- Los modelos econométricos de frontera estocástica*

*3.3- Los modelos de ajuste de riesgos*

*3.4- Problemas en la construcción de los ranking y diseminación de resultados de las ligas*

#### *4- Los Ranking multidimensionales*

*4.1- Métodos que utilizan pesos preestablecidos*

*4.2- Métodos que calculan los pesos de los distintos objetivos como resultado del análisis. El Análisis Envoltante de Datos (AED)*

*4.2.1- La formulación matemática de los modelos AED<sup>1</sup>*

*4.2.2- Extensiones del modelo AED original*

*4.2.3- Interpretación de resultados de los modelos AED. Un ejemplo sencillo*

*4.2.4- Aplicabilidad práctica de los modelos AED a centros asistenciales*

#### *5- Conclusiones*

<sup>1</sup> Este apartado tiene dificultad técnica. El lector no familiarizado con las matemáticas puede saltarlo sin perder el hilo del capítulo

Pero los centros asistenciales tienen objetivos múltiples, complementarios o no. Los métodos que integran múltiples objetivos en un indicador global son más complejos que los de criterio único. La elección del método debe estar en función de su uso. Cuanto más local sea el nivel de la gestión, mayor desagregación de indicadores se requiere, obligando a monitorizar los objetivos uno a uno. Los ranking basados en criterios múltiples buscan el «orden global» de un grupo de unidades (sean profesionales, servicios o centros asistenciales) requieren ponderaciones, establecidas «ex ante» por el investigador, o bien derivadas del propio análisis, como en el Análisis Envoltante de Datos (AED).

El análisis AED es un punto de encuentro entre académicos y gestores, para diseñar sistemas de incentivos, pagar a proveedores o hacer un seguimiento temporal del desarrollo organizativo, pero su aplicación en sanidad tiene problemas prácticos específicos.

## Introducción. Raking y Benchmarking

Hoy en día, se buscan Políticas Basadas en la Evidencia (PBE). El *benchmarking* consiste en comparar para mejorar emulando a los que lo hacen mejor. Se practica tanto a escala de *mesogestión* de centros como a escala mundial. Así, la Organización Mundial de la Salud (OMS) ha comparado los sistemas sanitarios en cuanto a los recursos que emplean y los resultados que consiguen. La OCDE define sistemas de indicadores sanitarios comparables internacionalmente. Las Naciones Unidas elaboran y difunden el Índice de *Desarrollo Humano* de los países del mundo, que combina varias dimensiones del bienestar —PIB, educación, salud.

El **benchmarking** se ha convertido en un nuevo paradigma de las políticas públicas **para mejorar por comparación**, y se aplica a países, organizaciones, centros asistenciales y profesionales individuales.

Este capítulo se dedica a las medidas de desempeño de los centros asistenciales que permiten comparaciones, y mejoras por comparación, de los centros y servicios.

### ¿Para qué sirven los métodos que miden el desempeño de los centros?

Estos métodos pueden servir para:

- diseñar sistemas de incentivos, para los gestores y para los clínicos
- facilitar las mejoras por comparación
- ajustar fórmulas de pago a proveedores basadas en el desempeño, el llamado *Payfor Performance* (P4P) entre otros
- proporcionar información relevante al ciudadano para elegir centro y servicio con criterios de calidad asistencial.

# 1- ¿Qué se entiende por desempeño?

## 1.1- Inputs, outputs y funciones de producción

En la **figura 1** se representan las relaciones entre inputs primarios, outputs intermedios y output final en la producción de salud.

La teoría microeconómica de la eficiencia productiva y funciones de producción (caja de herramientas de los economistas), relaciona directamente los recursos productivos o inputs primarios (tiempo de trabajo médico, camas, quirófanos) con los resultados o productos finales (enfermedades resueltas, mejoras de salud) pero, dada la dificultad de medir esos productos finales, las aplicaciones prácticas se limitan a evaluar las relaciones entre estructura y proceso, entre inputs primarios y actividades (estancias, altas, consultas), dando implícitamente por supuesta la calidad y adecuación del cuidado a unos estándares cuya definición se presupone.

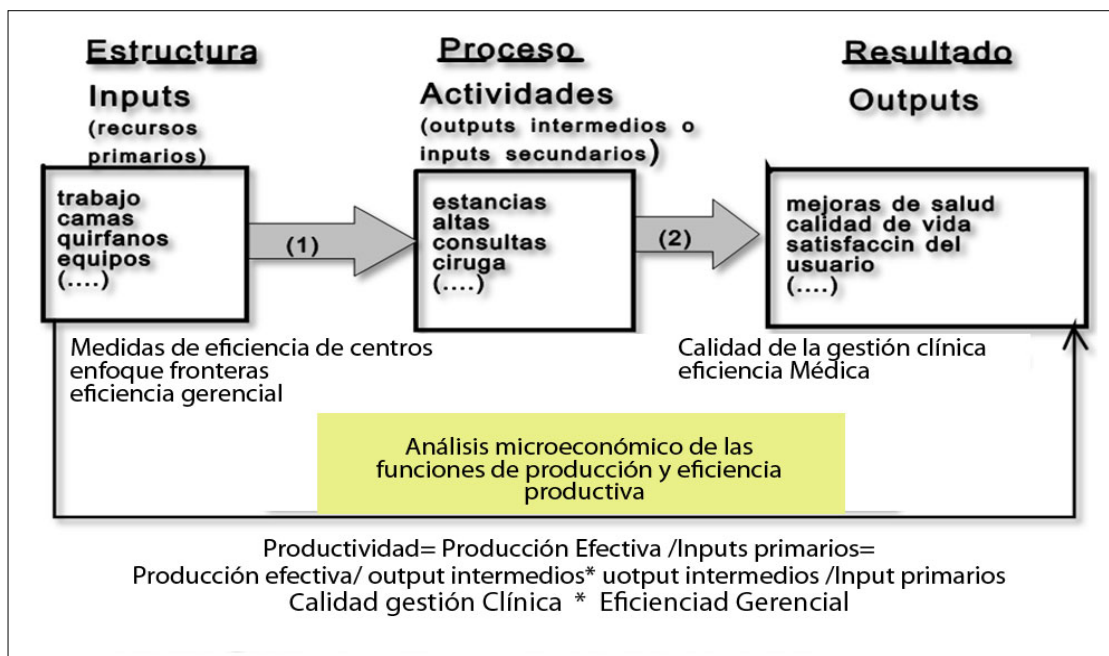


Figura 1: Relaciones entre inputs, outputs intermedios y outputs finales

Antes de seguir adelante conviene formalizar los conceptos de productividad, eficiencia y desempeño.

Los conceptos de productividad, eficiencia y desempeño están muy ligados.

**Productividad** es la ratio entre outputs obtenidos e inputs empleados en el proceso de producción.

**Eficiencia** compara en algún sentido los valores observados de outputs o de inputs con los óptimos. Por tanto, eficiencia siempre es un concepto relativo.

Los aumentos de productividad se pueden deber a mejoras de eficiencia, pero también a cambios tecnológicos, particularmente en un contexto dinámico, o a las condiciones del entorno en el que opera el centro. Eficiencia sería, pues, la productividad máxima alcanzable con una tecnología dada y en un entorno determinado. En cualquier caso, eficiencia equivale a 'éxito', a hacer las cosas bien y la ineficiencia resulta de un factor no observable y separable de los demás factores de producción, que es la gestión. Por eso, puede interpretarse que medir la eficiencia es medir la productividad marginal relativa de la función directiva.

La productividad como cociente entre producción efectiva (mejoras de salud) y los inputs empleados puede descomponerse en dos factores: la ratio entre producción efectiva y outputs intermedios y la ratio entre éstos y los inputs:

$$\begin{aligned}
 & \text{PRODUCTIVIDAD} \\
 & = \frac{\text{PRODUCCION EFECTIVA}}{\text{INPUTS PRIMARIOS}} \\
 & = \frac{\text{PRODUCCION EFECTIVA}}{\text{OUTPUTS INTERMEDIOS}} \cdot \frac{\text{OUTPUTS INTERMEDIOS}}{\text{INPUTS PRIMARIOS}} \\
 & = \text{CALIDAD GESTION CLINICA} \cdot \text{EFICIENCIA GERENCIAL}
 \end{aligned}$$

El primer factor se refiere a la **calidad de la gestión clínica** y el segundo a la **eficiencia gerencial**. Entendemos por **desempeño** la medida en que un centro cumple sus objetivos, que incluyen tanto la calidad clínica como la eficiencia de gestión.

*Productividad es la ratio entre outputs obtenidos e inputs empleados en el proceso de producción.*

*Eficiencia compara en algún sentido los valores observados de outputs o de inputs con los óptimos. Por tanto, eficiencia siempre es un concepto relativo.*

*Medir la eficiencia es medir la productividad marginal relativa de la función directiva.*

La evaluación de la eficiencia a nivel de gestión clínica sirve para detectar usos inapropiados de recursos. La identificación del uso inapropiado de recursos interesa a los clínicos (saber si están haciéndolo bien), pero también a los gerentes (saber si están despilfarrando recursos, haciendo pruebas diagnósticas innecesarias o generando estancias inapropiadas).

La definición de **eficiencia técnica**, ya clásica, se debe a Koopmans e implica dos condiciones: para aumentar la producción de cualquier output se necesita aumentar la cantidad de al menos un input o reducir la producción de al menos otro output y para reducir el empleo de cualquier input es preciso aumentar al menos otro input o reducir al menos un output. Eficiencia técnica equivale, cuando hay un solo output, a producir la mayor cantidad posible con unos recursos dados. En el concepto de eficiencia técnica subyace el máximo aprovechamiento de los recursos, no despilfarrar ningún input.

En 1957, Farrell propone dos medidas radiales alternativas de eficiencia técnica, la orientada a inputs y la orientada a outputs.

La distancia de Farrell **orientada a inputs**, DFI, es 1 menos la máxima reducción equiproporcional de inputs requerida para obtener un output dado ( $y$ ). La isocuanta  $L(y)$  es el conjunto de puntos en el espacio de factores tales que la eficiencia de Farrell orientada a inputs es la unidad (**figura 2**).

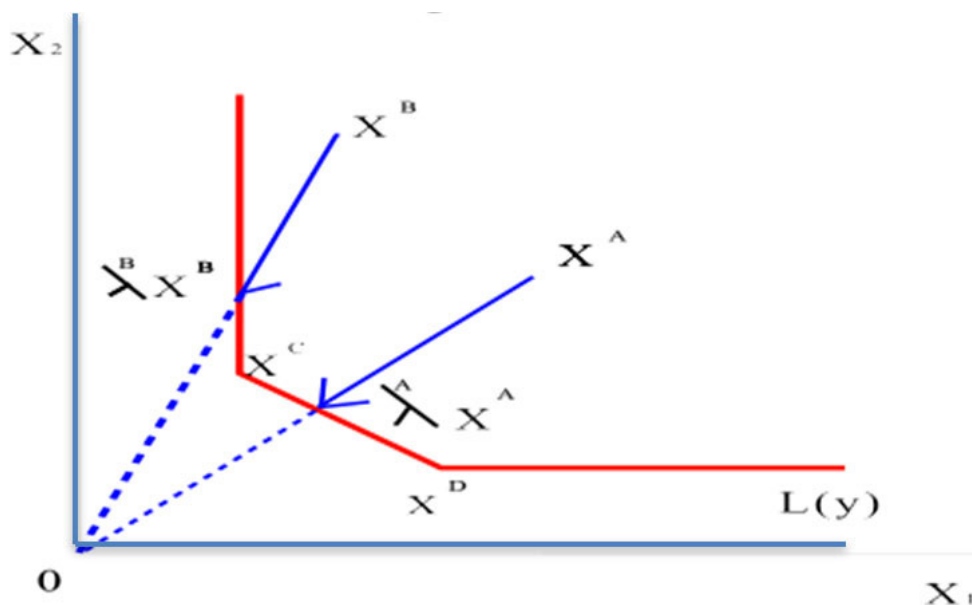


Figura 2. La distancia de Farrell orientada a inputs

*Entendemos por desempeño la medida en que un centro cumple sus objetivos, que incluyen tanto la calidad clínica como la eficiencia de gestión.*

De forma similar, se puede definir la medida de Farrell **orientada a outputs** (DFO), trabajando en el espacio de los productos (**figura 3**).

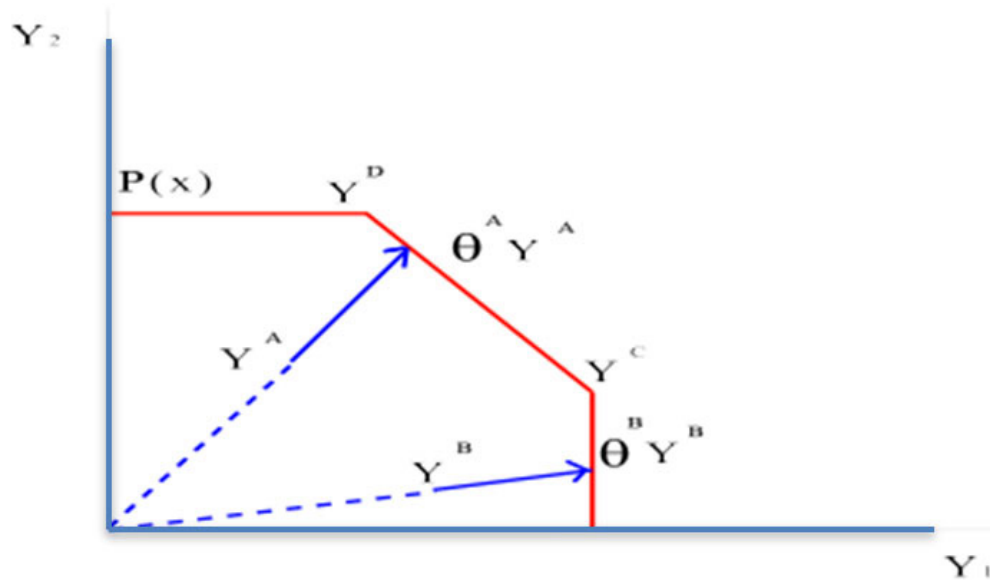


Figura 3. La distancia de Farrell orientada a outputs

Todo análisis de la eficiencia técnica admite esas dos orientaciones, a los inputs y a los outputs. Un centro asistencial es técnicamente eficiente, en el sentido de Koopmans, si ambos índices de Farrell de eficiencia, el orientado a inputs y el orientado a outputs, son la unidad.

La **eficiencia económica o asignativa** se refiere al grado de cumplimiento de un objetivo económico, sea éste minimizar costes o maximizar ingresos o beneficios. Requiere datos sobre cantidades y precios relativos –las medidas de ineficiencia técnica solo requieren datos de cantidades- y refleja los movimientos a lo largo de una isocuanta, como se muestra en la figura 4 para el caso del análisis orientado a inputs (la isocuanta es la curva que une los puntos de unidades que tienen el mismo nivel de producción)

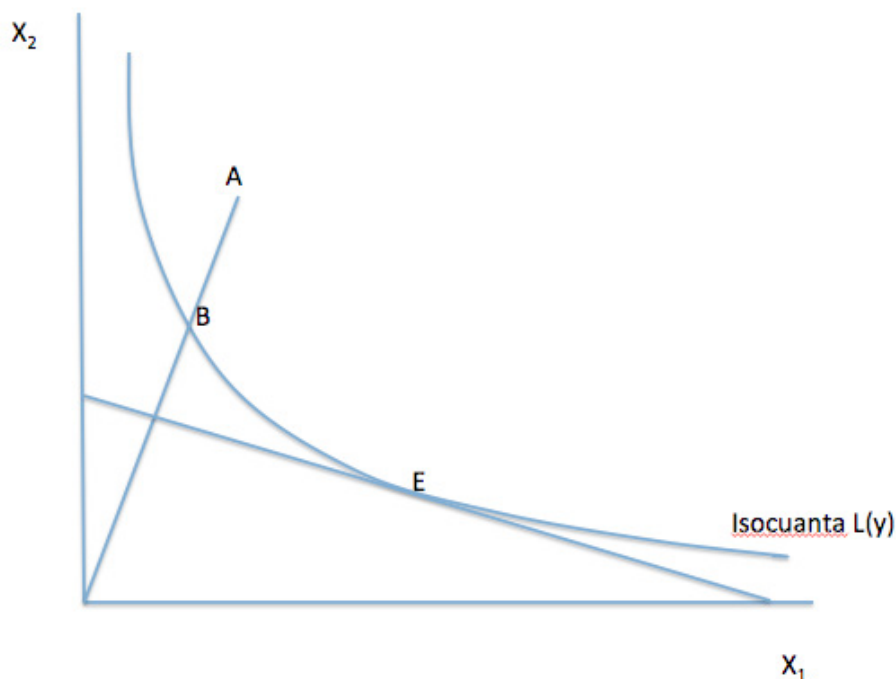


Figura 4. La eficiencia económica o asignativa

El centro A presenta ineficiencia técnica, pues no está sobre la isocuanta  $L(y)$ . A 'despilfarra' o emplea demasiados inputs para obtener la cantidad de output  $y$ . Presenta también ineficiencia asignativa porque emplea los dos inputs  $X_1$  y  $X_2$  en proporciones relativas subóptimas, dados sus precios relativos (pendiente de las rectas paralelas). El centro B es técnicamente eficiente, pero no lo es asignativamente. El comportamiento eficiente en costes viene representada por el centro E. **La eficiencia en costes implica eficiencia técnica y asignativa.**

## 1.2- Algunas experiencias a título ilustrativo

En los años 1990 se inició en el estado de **Nueva York** la recolección y publicación sistemática de indicadores de calidad de los servicios de **cirugía cardiaca** basados en datos de mortalidad, y de cirujanos individuales, con el fin de que los pacientes tuvieran información para elegir. Algunos de los servicios que quedaron en los últimos puestos, que tenían niveles de actividad bajos, cerraron.

En el **Reino Unido**, se elaboran y publican indicadores



comparativos de gestión, calidad asistencial y tiempos de espera entre áreas y centros sanitarios, bajo el epígrafe genérico **NHS outcomes and performance** (Resultados y desempeño del Sistema Nacional de Salud)<sup>1</sup>. Ese sistema de indicadores tiene la doble función de mejora de la gestión por comparación y de orientar a los pacientes sobre la calidad de los proveedores. Mas recientemente, se han empezado a publicar indicadores comparativos de desigualdades en salud y sus determinantes por zonas de Londres<sup>2</sup>.

En **España**, una empresa privada elabora desde 2000 un ranking de calidad y eficiencia de los hospitales, públicos y privados que voluntariamente participan, basado en la agregación de indicadores de calidad, procesos asistenciales y coste. Es el programa Top20<sup>3</sup>. También el Servicio Andaluz de Salud Pública indicadores comparativos de desempeño de sus centros<sup>4</sup>.

Desde la perspectiva de las políticas, son muy interesantes los experimentos de **pago por resultados (pay-for-performance, P4P)** a hospitales, que se inició experimentalmente en EEUU 2005. Su propósito es incentivar financieramente a los hospitales que dan asistencia de calidad a sus pacientes ingresados, dentro de los programas de aseguramiento público Medicare y Medicaid<sup>5</sup>.

1 <http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/Perfomancedataandstatistics/index.htm>

<http://www.hospitalcompare.hhs.gov/>

2 [http://www.lho.org.uk/LHO\\_Topics/national\\_lead\\_areas/marmot/marmotindicators.aspx?](http://www.lho.org.uk/LHO_Topics/national_lead_areas/marmot/marmotindicators.aspx?)

3 <http://www.commonwealthfund.org/Publications/Fund-Reports/2007/Apr/Pay-for-Performance-in-State-Medicaid-Programs--A-Survey-of-State-Medicaid-Directors-and-Programs.aspx>

[http://www.chcs.org/usr\\_doc/Physician\\_P4P\\_Guide.pdf](http://www.chcs.org/usr_doc/Physician_P4P_Guide.pdf)

4 [http://www.calidadsaludandalucia.es/es/apendice\\_tecnico.html](http://www.calidadsaludandalucia.es/es/apendice_tecnico.html)

<http://www.calidadsaludandalucia.es/es/index.html>

5 [http://www.cmio.net/index.php?option=com\\_articles&article=25595](http://www.cmio.net/index.php?option=com_articles&article=25595)

## 2.- Tipología de los métodos

Los métodos para medir el desempeño de los centros y obtener un ranking difieren según los objetivos, los datos y el modelo subyacente (grado de incertidumbre y forma de incorporar, en su caso, juicios de valor para ponderar diferentes objetivos).

Desde el punto de vista estadístico es relevante la diferenciación entre modelos deterministas y aleatorios, pero desde la perspectiva de la gestión de los centros y de las políticas es más interesante la clasificación de la **Figura 5**. En ella diferenciamos los ranking que corresponden a uno y a múltiples objetivos, y entre éstos separamos los métodos que ponderan los objetivos según criterio del que mide y los que obtienen pesos óptimos de cada uno de los objetivos -u outputs- como resultado del propio método.

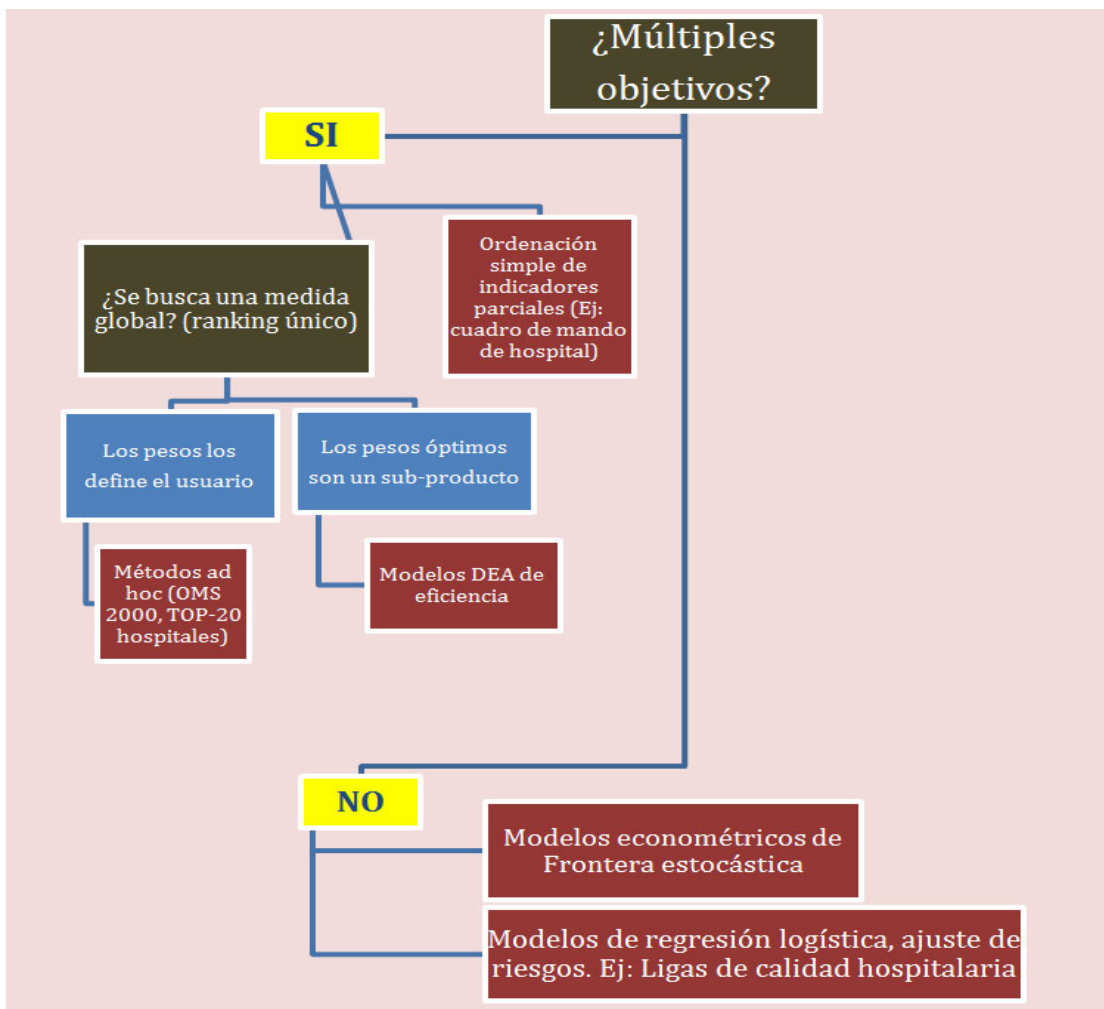


Figura 5: Tipología de los métodos de medida del desempeño en los centros sanitarios

### 3- Los ranking unidimensionales (basados en una medida única de "éxito" )

Cada vez con mayor frecuencia y trascendencia financiera se elaboran, publican y utilizan ordenaciones de unidades proveedoras de servicios públicos, por ejemplo hospitales, o centros de atención primaria, basados en un único criterio u objetivo de desempeño o de calidad.

La elección del criterio de calidad (mortalidad, complicaciones, reingresos) no es trivial. El indicador elegido debería ser fiable, válido, sensible, preciso, con interpretación clínica, útil tanto para la elección informada de hospital por los pacientes como para que los profesionales dispongan de estándares de buena práctica. Asimismo, debería ser fácil de obtener y difícil de manipular.

#### 3.1 Fases del proceso de construcción del ranking unidimensional de centros

El proceso de construcción de un ranking de indicador único de centros, que también podría servir para ordenar servicios o profesionales, discurre por las siguientes fases:

- 1) **Seleccionar** el indicador (por ejemplo, la tasa de mortalidad intrahospitalaria tras un infarto, o la satisfacción de los pacientes atendidos) y el periodo de referencia (por ejemplo, el último año)
- 2) **Calcular** el indicador bruto para cada centro, servicio o persona evaluada (i) con los datos registrados de sus  $n_i$  pacientes, a partir de una muestra representativa de los pacientes atendidos en el periodo de referencia, o de toda esa población
- 3) **Estandarizar** el indicador. El proceso de estandarización (ajuste por riesgo, por gravedad del paciente y por condiciones del entorno del hospital independientes de su práctica) requiere un modelo estadístico de ajuste que a su vez introduce incertidumbre adicional, por los errores de medida, de especificación y de estimación.

Los modelos estadísticos de ajuste más empleados (véanse apartados 3.2 y 3.3) son:

- a) los **modelos econométricos de frontera estocástica** de producción o de costes, que tratan de medir comparativamente la eficiencia de los centros, en el marco teórico de la teoría económica de la producción. En el apartado 3.2. definimos estos métodos con mayor profundidad
  - b) los **modelos de ajuste de riesgos**, basados en **regresión logística** o en otros métodos similares, que estiman la calidad (por ejemplo, mortalidad hospitalaria –complicaciones, reingresos,...- esperada de cada centro, ajustando por severidad de los casos tratados, y en su caso por otros factores no controlables de entorno). En el apartado 3.3. profundizamos en estos métodos
- 4) **Ordenar** los centros, servicios o personas según el indicador estandarizado, calcular los intervalos de confianza y contrastar la significación de las diferencias.

Cada una de estas fases es susceptible de problemas estadísticos y prácticos que pueden invalidar los resultados (véase apartado 3.4).

*El proceso de construcción del ranking tiene las fases de Seleccionar (indicadores), Calcular, Estandarizar, y Ordenar (los centros)*

### **3.2- Los modelos econométricos de frontera estocástica**

Estiman el grado de ineficiencia productiva de un conjunto de unidades de producción que en nuestro caso serán centros asistenciales o servicios. Estiman la ineficiencia técnica y/o asignativa, con fronteras de producción o de costes, suponiendo que la propia frontera es estocástica. Son modelos con un "error compuesto" de dos componentes. Uno de ellos forma parte de la frontera y representa los errores de medida, omisión de variables y presencia de acontecimientos no predecibles y fuera de control que afectan a la variable dependiente (actividad o coste). El segundo componente, asimétrico, (positivo en modelos de fronteras de costes, y negativo en modelos de fronteras de producción), capta la ineficiencia.

Un modelo de frontera estocástica de **producción** para un único output (y) de la unidad asistencial i-ésima se formularía así:

$$\log(Y_i) = x_i \beta + z_i \gamma + v_i + u_i$$

$$\log(Y_i) = x_i \beta + z_i \gamma + v_i + u_i$$

Donde  $x'$  es un vector de  $k$  factores de producción, como número de médicos, presupuesto, quirófanos (o una función de ellos, por ejemplo, el logaritmo),  $z'$  es un vector de variables de entorno no controlables que influyen en la producción (la morbilidad o la estructura etaria de la población atendida, por ejemplo),  $v$  es un error aleatorio que se supone distribuido con una función de distribución continua de media cero independiente de las  $x$  y de las  $z$ , y  $u$  es la ineficiencia, variable aleatoria negativa o nula, distribuida independientemente de las  $x$  y de las  $z$ . El modelo de frontera de costes se formularía de forma similar, salvo que  $u_i \geq 0$ . Estos modelos admiten diversas formas funcionales.

En el modelo para hospitales, el output  $y$  de la función de producción podría ser el número de UPAs (Unidades Ponderadas Asistenciales), o el número de pacientes tratados, por ejemplo. Obviamente, un problema de estos modelos es que no consideran que los centros asistenciales son multiproducto (consultas, intervenciones quirúrgicas, ingresos,...).

Aunque se han sofisticado en varias direcciones, los modelos de frontera estocástica no están todavía listos para ser empleados con propósitos prácticos de medir la eficiencia con fines remunerativos. Su atractivo, hasta ahora, es más académico que profesional. En su aplicación a sanidad, una parte del problema es que están desvinculados de los resultados clínicos y de la calidad asistencial. Los ajustes por calidad son un problema común en la evaluación de servicios públicos de múltiples outputs, donde

el mercado no emite señales porque no se comercializan (las universidades, o la policía serían equiparables en este sentido a los hospitales).

---

### 3.3- Los modelos de ajuste de riesgos

---

Otros modelos se emplean para medir la calidad clínica de los hospitales, para emitir señales de calidad y orientar la elección de los pacientes, en un contexto de política sanitaria de autorregulación por el mercado. Se basan en ajustar modelos de **regresión logística** que explican un evento (complicación, reingreso, o muerte) en función del riesgo basal del paciente, con tantas covariables como sean necesarias (GRD, edad, comorbilidad,...). La variable dependiente o variable a explicar en estos modelos es una dicotómica que vale 0 o 1. Por ejemplo, vale 1 si el paciente reingresa por el mismo diagnóstico o por un diagnóstico que pueda estar asociado al ingreso original por causa de una complicación<sup>6</sup>, en el mes siguiente al alta. La parte de la varianza no explicada por ninguna de las variables de ajuste (que son las variables "explicativas" del modelo) cuantifica los resultados adversos que no pueden ser justificados por el riesgo basal del paciente; este "residuo" estadístico, sería a priori atribuible a otros factores, y medirían con mayor precisión la parte de la calidad o el desempeño que depende del centro, servicio o profesional. Por ello, con estos residuos promedio de los pacientes de cada hospital se calcula la puntuación del mismo, en comparación con el conjunto de hospitales.

---

### 3.4- Problemas en la construcción de los ranking y diseminación de resultados de las ligas

---

1. ¿Qué queremos comparar? Método de cálculo de los intervalos de confianza y significación de las diferencias
  - Según el objetivo de las comparaciones, habrá que elegir el método adecuado de contraste de la significación de las diferencias entre centros. Aunque se trata de una cuestión técnica en la que no profundizaremos, conviene saber que para hacer comparaciones múltiples

---

<sup>6</sup> El algoritmo para determinar en que pacientes la variable dependiente vale 1 es complejo y de su adecuación y validez depende todo lo demás. Por ejemplo, hay que excluir diagnósticos que impliquen ingresos sucesivos (por ejemplo, cáncer que requiere ciclos de quimioterapia con ingreso hospitalario).

debemos construir intervalos de confianza múltiples o simultáneos, para lo cual puede usarse un método basado en la desigualdad de Bonferroni.

## 2. Problemas de tamaño. Efectos tamaño e incertidumbre

Puesto que los indicadores medidos son al fin y al cabo extracciones aleatorias de una distribución de probabilidad, hay errores de muestreo que varían entre centros según su tamaño. Muchos ranking publicados no informan al respecto.

Veamos un ejemplo ilustrativo sencillo. Sean dos centros, uno es un pequeño hospital con 40 intervenciones anuales de cirugía cardíaca y el otro es un complejo hospitalario que hace 1.000 intervenciones al año. Para estimar la tasa de complicaciones (o de mortalidad, o reingresos) se calcula la frecuencia muestral (es decir, la proporción de pacientes intervenidos que sufrieron el suceso adverso. Si los pacientes son homogéneos, no hay sesgo de selección y ambos hospitales tienen una probabilidad idéntica de que ocurra el suceso adverso (igual calidad). Los números de sucesos adversos siguen sendas distribuciones binomiales con parámetros  $(n,p)$  donde  $n$  es 40 para el hospital pequeño y 1.000 para el grande, y  $p$  es idéntica (la probabilidad del suceso adverso).

Dados  $n$  y  $p$  podemos calcular los intervalos simétricos respecto a  $p$  que contienen una proporción determinada, por ejemplo el 95%, de la masa de probabilidad de la distribución respectiva. Hemos representado en la figura 5 dichos límites ( $p_1$  y  $p_2$ ) para ambos hospitales, para una probabilidad del 95% y niveles de calidad (tasas de efectos adversos,  $p$ ) entre el 5% y el 25%. Como puede observarse, los intervalos son mucho más amplios para el pequeño hospital que para el grande. Además, las distancias se acrecientan para sucesos adversos más probables (al aumentar  $p$ ).

Como vemos en la **Figura 6**, los IC del 95% dependen del número de casos tratados y de la probabilidad de éxito y de fracaso ( $p_i$  y  $1-p_i$ ). Cuanto menor sea el tamaño o nivel de actividad de un hospital, menor será la potencia de los

contrastes sobre su calidad. Es decir, hay mas incertidumbre sobre su calidad real. Los hospitales pequeños tienen intervalos de confianza amplios, compatibles con un amplio rango de calidad asistencial. En la práctica, pues, tendrán mayor probabilidad de ocupar los puestos extremos del ranking. Si las listas no informan sobre intervalos de confianza, están contaminadas por el efecto tamaño y las comparaciones pueden ser engañosas.

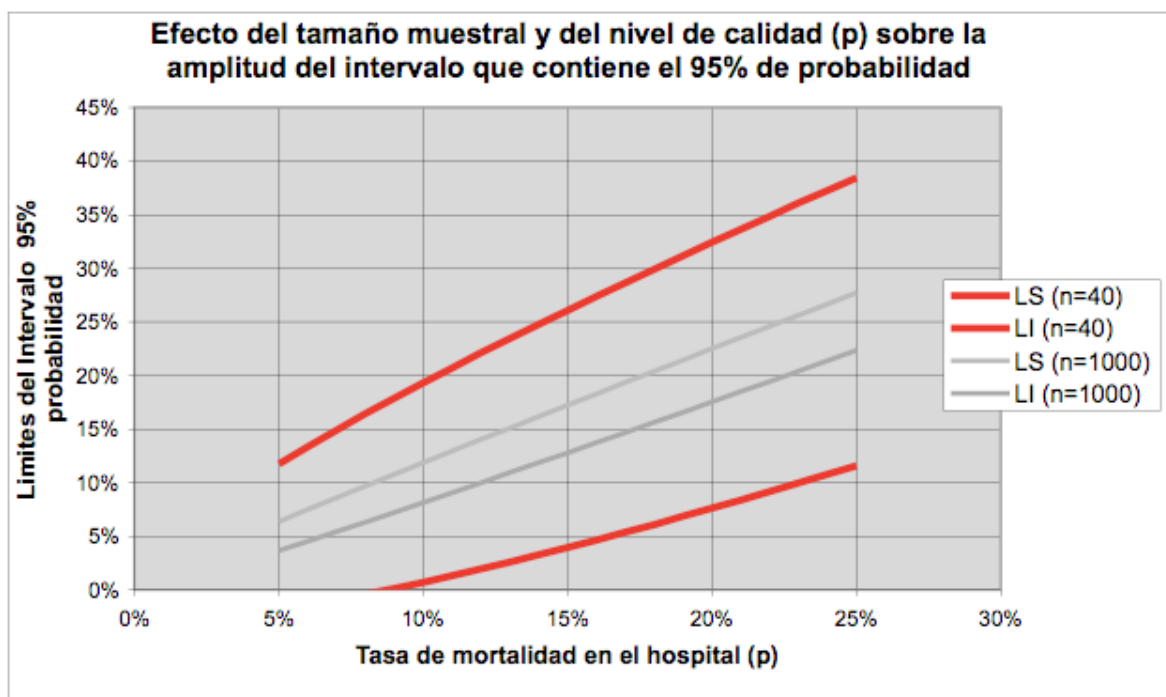


Figura 6: Tamaño muestral y nivel de calidad en relación al intervalo de confianza (Nota: los límites se han calculado aproximando la distribución binomial a la Normal)

Pero no solo hay contaminación por el efecto tamaño. Es más grave el **sesgo de selección** y la heterogeneidad no observable entre pacientes. Por ejemplo, la mortalidad de los pacientes ingresados con infarto de miocardio puede depender más del estado de salud en que llegan al hospital que de la calidad de la atención sanitaria que éste les presta. Puesto que hay sesgo de selección (los mejores hospitales reciben a los pacientes más graves), si no se ajusta bien por gravedad se encuentra mayor mortalidad asociada a los mejores hospitales.



### 3. *Problemas de ajuste y "comparaciones ajustadas"*

El ajuste por riesgo es tan esencial, que un ranking que no lo haga no sólo es inútil, sino engañoso y contraproducente. En efecto, puede haber un problema de endogeneidad. La tasa de mortalidad no refleja realmente la calidad del servicio sino su prudencia, al admitir solo a los casos mas leves y con mayor probabilidad de éxito. Un problema añadido de las ligas públicas mal ajustadas por riesgo es que nadie acepta operar a los pacientes mas viejos o de mayor riesgo por evitar bajar en la liga.

La "depuración" de los datos brutos para obtener un indicador estandarizado o ajustado por la gravedad inicial de los pacientes y otras condiciones independientes del hospital (ajuste por riesgo) requiere un modelo de ajuste, siendo la regresión logística uno de los mas empleados. En cualquier caso, la variabilidad entre hospitales se estrecha tras la estandarización. El indicador estandarizado, se basa, en definitiva, en los residuos de esa regresión, que aproximan la sobre-mortalidad no justificable por las condiciones en que opera el hospital.

### 4. *Problemas de datos*

La precariedad de los datos y los errores en las variables pueden invalidar los resultados. Estos errores pueden deberse a problemas de los sistemas de información, pero también ocurre que los propios hospitales manipulan a su favor la información que difunden, y optimizan el comportamiento del indicador de la calidad y no la calidad misma.

### 5. *Problemas de difusión de los resultados*

Siendo la metodología compleja, los resultados que se difunden se reducen a una tabla. Sólo en letra pequeña se aclaran las limitaciones metodológicas. La simplificación induce a interpretaciones erróneas sobre la significatividad de las diferencias en los puestos de la clasificación. Convendría publicar las tasas brutas y las ajustadas, el tamaño (número de casos) y los intervalos de confianza en torno a la tasa media global común.

## **4- Los Ranking multidimensionales**

---

Tanto las políticas como la gestión de centros asistenciales tienen objetivos múltiples, complementarios o no. La cuestión es si ponderar o no ponderar. Cuanto más local sea el nivel de gestión, se requiere mayor desagregación de indicadores, y monitorizar los objetivos uno a uno. Para asignar fondos entre unidades con las mismas reglas del juego, el regulador central requerirá una medida sintética de desempeño que compense las pérdidas respecto a la media de unas dimensiones con las ganancias de otras dimensiones. Cuanto más centralizado es el organismo que necesita el ranking, más necesidad suele haber de sintetizar las múltiples dimensiones del éxito en un único indicador global de "desempeño" o eficiencia organizativa y calidad asistencial. El gerente de un hospital, por ejemplo, necesita en su cuadro de mando indicadores separados de productividad y actividad en cada uno de los servicios y unidades, para tomar medidas específicas que resuelvan los problemas. Necesitará conocer la lista de espera de cada servicio y prueba diagnóstica. En cambio la dirección regional del sistema de salud y la autoridad sanitaria central demandarán un indicador global de eficiencia.

Los métodos que buscan el "orden global" perfecto de un grupo de unidades (sean países, hospitales o problemas de salud) son complejos y requieren ponderaciones de los diferentes outputs o dimensiones del desempeño. Esas ponderaciones pueden ser establecidas "ex ante" por el investigador (apartado 4.1), o bien resultar del propio análisis (apartado 4.2).

### **4.1- Métodos que utilizan pesos preestablecidos**

---

El ranking de sistemas de salud del mundo del informe 2000 de la OMS(WHO, 2001) resultó de la ponderación ad hoc (con criterios OMS) de las dimensiones que, según un panel de expertos, definen el desempeño de los sistemas de salud. Tiene el mérito del pionero, cuya estela han seguido muchos trabajos, organismos e investigadores.

En septiembre de 2001, el NHS británico empezó a publicar, entre otros, un ranking de los consorcios que proveen asistencia hospitalaria aguda, clasificándolos con estrellas de excelencia, entre cero y tres según un amplio conjunto de indicadores de cumplimiento de objetivos clínicos, de gestión, de calidad, de

tiempos de espera y otros que interesan al paciente. En 2004 el Ministerio de Salud encargó a un grupo de investigación del Centre for Health Economics de la Universidad de York y del National Institute for Economic and Social Research el diseño de una nueva metodología para medir la productividad y resultados globales del NHS. El informe, publicado en 2005 (Dawson et al., 2005), propone un índice global de output del NHS ajustado por calidad (el “*valueweighted output index*”).

El índice requiere datos de actividad y de resultados (en términos de de salud y satisfacción de los pacientes, por ejemplo tiempos de espera, que afectan a la utilidad, o la incertidumbre asociada a esa espera. Además, es preciso tener datos de la valoración social de cada uno de esos resultados ( $\Pi$ ), que son las ponderaciones incorporadas en la fórmula.

El informe británico abrió nuevos horizontes y cambió perspectivas, sobre todo porque se centra en el paciente, y la calidad se define en función de las dimensiones de los resultados que los pacientes valoran.

Una experiencia de benchmarking de hospitales en España es el Top20, que incorpora varios indicadores de calidad clínica, tecnología y eficiencia de costes. Los hospitales son comparados con los de su grupo, previamente se construyen grupos homogéneos según tamaño, tecnología y propiedad pública o privada. En los últimos años, el programa Top20 se ha extendido a la elaboración de ranking específicos por servicios o especialidades, y a la atención primaria.

Lo más discutible de los métodos que ponderan con pesos ad hoc es la propia elección de esos pesos para construir un índice global de desempeño. Por ejemplo, cómo conciliar la excelencia global con malos indicadores en alguno de los indicadores parciales.

---

#### **4.2- Métodos que calculan los pesos de los distintos objetivos como resultado del análisis. El Análisis Envolverte de Datos (AED)**

---

Entre los métodos que evalúan la eficiencia organizativa global de unidades prestadoras de servicios homogéneos no comercializados, donde no hay precio de mercado que oriente sobre el “valor” de los bienes o servicios, el **Análisis Envolverte de Datos, AED (Data Envelopment Analysis, DEA)** tiene

gran aceptación y es un punto de encuentro entre académicos y gestores<sup>7</sup>.

Los resultados sirven para diseñar incentivos, pagar a proveedores o hacer un seguimiento temporal del desarrollo organizativo. Se aplican sobre todo a servicios públicos que producen múltiples outputs sin precio de mercado que refleje su valoración relativa: educación, sanidad.

En sanidad, se han aplicado modelos DEA para tantear la posibilidad de medir la eficiencia comparativa de Autoridades Sanitarias, de hospitales y de Equipos de Atención Primaria, también en España.

Generalmente, se plantean como métodos sustitutivos de los modelos econométricos de frontera estocástica. Como ellos, todavía tendrán que resolver algunas cuestiones metodológicas pendientes y someterse al tribunal de los hechos para demostrar que son instrumentos útiles, precisos y robustos, para orientar el reparto de fondos entre centros y servicios.

Cómo gestionar científicamente la incertidumbre, cómo ajustar por factores "inevitables" condicionantes del entorno fuera de control de los gestores; o cómo tener en cuenta la dinámica de la eficiencia son algunas de las cuestiones pendientes.

---

#### 4.2.1- La formulación matemática de los modelos AED<sup>8</sup>

---

Los modelos AED aplican las **medidas de eficiencia de Farrell** mediante la resolución de problemas de **programación matemática** de tipo determinista. Para cada centro asistencial o servicio clínico, que de forma genérica denominaremos Unidad de Toma de Decisiones (**UTD**), determinan la '**frontera de la mejor práctica**', o grupo de UTD que, combinadas en una UTD compuesta y ficticia, logran resultados mejores. Son métodos no paramétricos porque no requieren especificar una determinada forma de la función de transformación de inputs en outputs.

A partir del modelo AED original, llamado CCR por las iniciales de sus autores (Charnes, Cooper y Rhodest, 1978) se han propuesto variantes progresivamente sofisticadas. Estos modelos de frontera

<sup>7</sup> En la página web <http://www.deazone.com/> hay recursos sobre los modelos AED: cursos, software, bibliografía, etc. Incluye aplicaciones a hospitales y otros centros: <http://deazone.com/en/download> que pueden bajarse de internet

<sup>8</sup> Este apartado tiene dificultad técnica. El lector no familiarizado con las matemáticas puede saltarlo sin perder el hilo del capítulo

determinista suponen que la frontera de producción o de costes de cada empresa carece de elementos aleatorios. Se formulan como problemas de programación matemática, generalmente lineal, e incorporan ,múltiples inputs y múltiples outputs.

Hay diferentes tipos de modelos DEA, que se han desarrollado a partir del original de Charnes, Cooper y Rhodes (CCR), presentando todos ellos dos versiones, la **orientada a inputs** y la **orientada a outputs**. El modelo CCR original orientado a outputs se formula como uno de programación fraccional cuya función objetivo es **maximizar la eficiencia técnica**, definida mediante el cociente entre la suma ponderada de los outputs y la suma ponderada de los inputs de la UTD. Supongamos que intervienen  $m$  inputs  $x_j$  ( $j=1, \dots, m$ ) y se producen  $k$  outputs  $y_r$  ( $r=1, \dots, k$ ). El objetivo del modelo radica precisamente en calcular los pesos óptimos  $v_j$  y  $u_r$  de los inputs y outputs respectivamente para cada UTD. Los pesos, llamados algunas veces **multiplicadores virtuales** varían entre UTD, se supone que respondiendo a sus diferentes sistemas de valores. Para cada una de las  $n$  UTD de la muestra ( $i_0=1, \dots, n$ ) se resuelve el siguiente problema:

$$\text{Max } h_0 = \frac{\sum_{r=1}^k u_r y_{ri_0}}{\sum_{j=1}^m v_j x_{ji_0}} \quad [1]$$

sujeto a:

$$\frac{\sum_{r=1}^k u_r y_{ri}}{\sum_{j=1}^m v_j x_{ji}} \leq 1, \quad i = 1, 2, \dots, n$$

$$u_r, v_j \geq 0, \quad \forall r, j$$

donde la función objetivo maximiza la eficiencia de la UTD  $i_0$ , restringida a que cada UTD en la muestra tiene, como máximo, eficiencia unidad, y que los pesos han de ser no negativos.

Para facilitar su resolución, el modelo [1] se transforma en uno de programación lineal, fijando arbitrariamente la suma ponderada de los inputs, que son el denominador de la función objetivo (por ejemplo, asignándole el valor 1) y transformando la función objetivo, (eficiencia de la UTD  $i_0$ ), en la suma ponderada de los outputs que obtiene. La versión primal del modelo resultante, cuya interpretación es similar a [1], es la siguiente:

$$\begin{aligned} \text{Max } h_0 &= \sum_{r=1}^k u_r y_{ri_0} \\ \text{Sujeto a:} \\ \sum_{j=1}^m v_j x_{ji_0} &= 1 \quad [2] \\ \sum_{r=1}^k u_r y_{ri} - \sum_{j=1}^m v_j x_{ji} &\leq 0, i = 1, \dots, n \\ u_r, v_j &\geq 0 \quad \forall r, j \end{aligned}$$

Este modelo tiene  $m+k$  variables (los **pesos** asignados por la unidad  $i_0$  a los **m inputs** y a los **k outputs**), y tantas restricciones como UTD hay en la muestra ( $n$ ) más una (que fija arbitrariamente la suma ponderada de los inputs igual a 1), aparte de las restricciones que imponen el signo de los pesos, pues no pueden ser negativos.

La solución de los  $n$  modelos de programación lineal [2] proporciona la **eficiencia relativa** de las  $n$  UTD. Son eficientes aquellas cuyo valor óptimo de la función objetivo es la unidad. Para cada UTD ineficiente, la solución del **dual** ( las restricciones con holgura en [2]) identifica el subconjunto de unidades eficientes de referencia, o **unidades envolventes**, las cuales tienen sus mismos pesos  $u_r$  y  $v_j$ . El valor de  $h^*_0$  es el índice de eficiencia de la UTD  $i_0$ , siendo  $h^*_0$  el valor óptimo de  $h_0$ .

Los valores óptimos de los  $u_r$  y los  $v_j$  en la solución primal [2] permiten obtener las **tasas marginales de transformación de outputs y de sustitución de inputs** respectivamente para cada UTD. Así,  $u_{r'}/u_r$  es la tasa marginal de transformación del output  $r$  por el output  $r'$ , indicando cuántas unidades de  $r'$  hay que aumentar si se reduce en una unidad la producción del output  $r$ .

El problema dual equivalente a [2] impone restricciones sobre el uso mínimo de inputs y sobre la obtención máxima de outputs. Sus variables corresponden a las restricciones del primal, y permiten construir una **"unidad compuesta"**, que al menos iguale en eficiencia a  $i_0$ , distinguiendo las UTD eficientes de referencia a "imitar". Es la "envolvente" de la UTD que se evalúa.

Hoy en día disponemos de un amplio y variado menú de opciones DEA. La elección depende de las características del problema a resolver; en consecuencia, una aplicación mecanicista es tan inadecuada y falta de justificación como en la modelización econométrica.

Las extensiones y avances recientes del análisis envolvente de datos apuntan en dos sentidos: aleatoriedad y dinamicidad. Otras extensiones permiten introducir variables categóricas, variables fuera de control de la UTD o restricciones sobre los pesos.

---

#### 4.2.2- Extensiones del modelo AED original

---

El modelo base CCR, que hemos formulado en versión orientada a outputs y que puede escribirse sin dificultad, alternativamente, orientado a inputs, impone tres restricciones sobre la tecnología frontera: **rendimientos constantes a escala, disponibilidad fuerte de inputs y outputs y convexidad** del conjunto de combinaciones factibles input-output. A lo largo de las últimas décadas han ido surgiendo modificaciones al modelo CCR original que relajan estas hipótesis.

Concretamente, el modelo conocido por **BCC** y formulado en 1984 por Banker, Charnes y Cooper admite que cada UTD puede operar con **rendimientos a escala no constantes** y permite evaluar la **escala óptima** de operación de cada UTD de la muestra, así como separar la parte de la ineficiencia debida estrictamente a problemas de gestión de la que se podría atribuir a los problemas de dimensión. Los rendimientos a escala resultantes de un análisis de envolvimiento de datos son locales en un punto de la superficie eficiente de producción, mientras que los modelos de frontera estocástica estiman rendimientos medios.

La hipótesis de **disponibilidad fuerte**, de inputs y/o outputs, puede relajarse añadiendo restricciones. La disponibilidad fuerte **de inputs** indica que se puede disponer libremente de excedentes de algunos inputs, en cantidades ilimitadas (trazos horizontal y vertical en la figura 2). La disponibilidad fuerte **de outputs** significa que pueden existir cantidades de output cuya producción no implica costes adicionales (trazos horizontal y vertical de la figura 3). Los modelos DEA permiten 'contrastar' la existencia de estos fenómenos, comparando los resultados que proporciona el análisis no restringido con el restringido a una disponibilidad débil.

También hay versiones modificadas que relajan la hipótesis de **convexidad** de la tecnología de transformación de inputs en outputs. En la versión original CCR, se compara cada UTD con una combinación lineal de otras UTD, combinación que la supera en eficiencia. Puede que en la práctica esa UTD compuesta, ficticia no sea factible. Añadiendo una restricción adicional se impone que, para ser declarada ineficiente, una UTD ha de ser dominada por otra UTD real y no por una supuesta UTD 'híbrida'. Son las llamadas fronteras **FDH ('free disposalhull')**, en las que subyace una noción de dominación más fuerte que en las fronteras AED.

Un problema adicional de las versiones AED que aplican las **medidas radiales** de Farrell es que representan, como hemos visto, únicamente condiciones necesarias para la eficiencia tal como es definida por Koopmans. Pueden encontrarse en la literatura versiones DEA, que incorporan medidas no radiales de eficiencia.

Los AED **estocásticos** tratan de superar el carácter determinista propio de estos modelos, que confunden ruido estadístico con ineficiencia, admitiendo cierta probabilidad de que las restricciones no se cumplan. Por el momento, se trata de una alternativa más teórica que práctica porque demanda mucha información (vectores de esperanzas y matrices de covarianzas de inputs y outputs se suponen conocidos), requiere hipótesis estocásticas de dudoso realismo (normalidad multivariante de todos los datos del problema) y obliga a resolver problemas computacionalmente complejos, de programación no lineal.



Los AED **dinámicos**, por el contrario, se han desarrollado espectacularmente en años recientes como alternativa metodológicamente aceptable y empíricamente viable para analizar cambios de la eficiencia a lo largo del tiempo. El análisis de ventanas consiste en realizar para una muestra de  $n$  UTD y  $T$  periodos, una secuencia de  $T-s+1$  análisis DEA abarcando cada uno  $s$  períodos consecutivos, con objeto de estimar las tendencias en la eficiencia y de hacer diagnóstico de outliers (valores extremos), que es uno de los grandes retos pendientes de las fronteras no paramétricas. Es una estrategia particularmente útil cuando se dispone de series longitudinales de las  $n$  UTD relativamente largas y queremos modelizar un número relativamente grande de inputs y outputs, relativo a  $n$ . De esta forma, en cada análisis el tamaño muestral se multiplica por  $s$ , 'ganando' grados de libertad.

### 4.2.3- Interpretación de resultados de los modelos AED. Un ejemplo sencillo

Sean 3 hospitales que emplean dos inputs, en cantidades a las que llamamos  $I_1$  e  $I_2$ , para producir tres tipos de outputs ( $O_1$ ,  $O_2$  y  $O_3$ ). En este ejemplo simplificado hay muy pocos centros, inputs y outputs, en los análisis reales habrá más inputs: camas, presupuesto de capítulo 2, enfermeros, quirófanos,... y de outputs: cirugía ambulatoria, hospitalizaciones, consultas, satisfacción, etc.

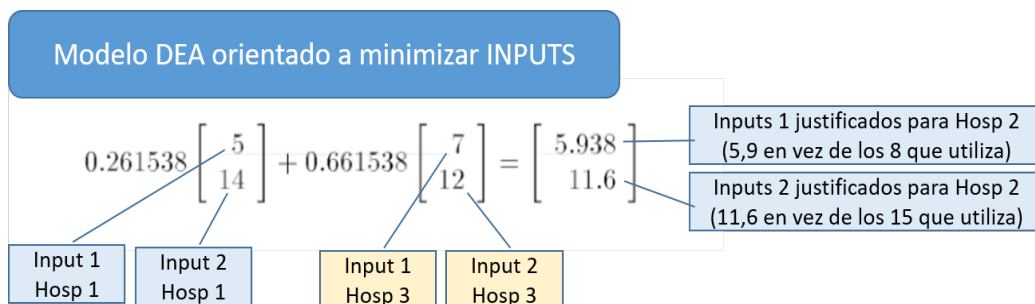
La tabla de datos es la siguiente:

Hospital	Input 1	Input 2	Output 1	Output 2	Output 3
1	5	14	9	4	16
2	8	15	5	7	10
3	7	12	4	9	13

La aplicación del modelo básico CCR con rendimientos a escala constantes da como resultado que los hospitales 1 y 3 son eficientes. El hospital 2 es ineficiente, siendo su ratio de eficiencia igual a **0.77333**.

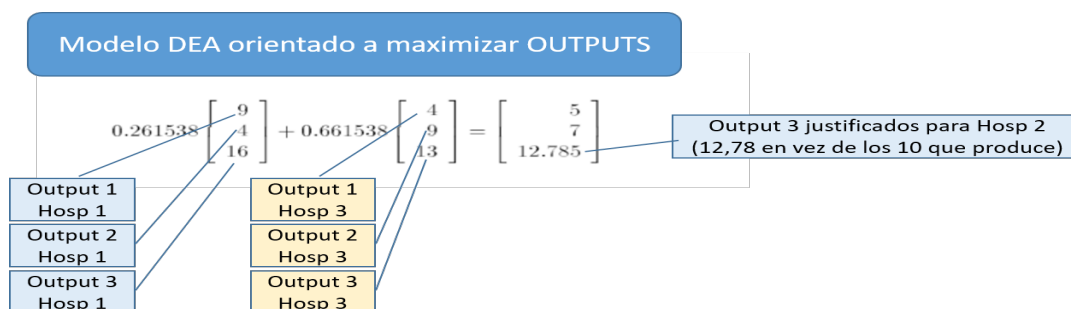
Una combinación lineal de los hospitales 1 y 3 "mejora" el desempeño del hospital 2, en el sentido de que necesitaría

menos cantidades de inputs de las que de hecho emplea para producir las mismas cantidades de outputs. Los hospitales 1 y 3 definen la frontera eficiente de la mejor práctica para el hospital 2, su envolvente esa combinación lineal de hospitales eficientes a los que debería imitar el hospital 2.



El resultado del modelo indica que el hospital 2 podría seguir produciendo los mismos niveles de sus tres outputs empleando 5.9368 unidades del I1 (utiliza 8) y 11.6 unidades de I2 (utiliza 15):

Visto desde la óptica de los outputs, el hospital 2 podría producir más cantidades del tercer output (12.785 en vez de 10, que es lo que de hecho obtiene) con los inputs de que dispone si fuera tan eficiente como los otros dos. No podría, sin embargo, producir mas de O1 ni de O2:



#### 4.2.4- Aplicabilidad práctica de los modelos AED a centros asistenciales

Hay centenares de aplicaciones de modelos AED para medir la eficiencia de hospitales y de centros de atención primaria. Una revisión puede encontrarse en Puig (2004). Sin embargo, hay problemas a los que dedicamos este apartado.

### **Problema 1. Precios sombra poco verosímiles y UTD con eficiencia espuria.**

- El modelo obtiene, para cada centro, los pesos que lo hacen aparecer como más eficiente. Pero el argumento de que los diferentes Centros ponderan de distinta forma sus recursos y sus servicios porque poseen "**sistemas de valores**" diferentes resulta bastante etéreo. Sería más razonable y consistentes con la teoría económica considerar que esos pesos representan los **precios sombra** de los outputs y de los factores de producción. Pero muchas veces, los resultados obtenidos para algunas UTD no parecen precios sombra realistas.
- Surge entonces la duda respecto a si una UTD que el modelo considera eficiente lo es en realidad, o por el contrario se trata sólo de una eficiencia aparente o espuria, derivada de la especial estructura de sus pesos. Un centro puede resultar eficiente simplemente porque es tan diferente de las demás en la composición de sus inputs o en la cartera de sus productos que no existen otras en la muestra con las que se le pueda comparar. Al ser imposible construir una unidad compuesta, por combinación lineal de algunas existentes, que lo haga mejor, la UTD *ectópica* sale del DEA con el cartel de *eficiente*.
- No obstante, dando la vuelta al argumento, esa limitación es también la gran virtud de los modelos DEA ; porque si una UTD no es eficiente ni siquiera con la mejor estructura posible de pesos, su falta de eficiencia queda fuera de toda duda.
- En síntesis. el modelo descubre a las unidades ineficientes, y cuantifica el porcentaje máximo en que se pueden reducir todos sus inputs sin disminuir la producción, por comparación con otras unidades. Sin embargo, no siempre identifica las unidades eficientes.
- Hay dos tipos de **soluciones**: imponer restricciones adicionales sobre la importancia relativa de los inputs y outputs y excluir de la muestra las UTD cuyos pesos resultan claramente no razonables.

## Problema 2. Sensibilidad de los resultados a la presencia de outliers.

- Este problema surge como consecuencia de su carácter determinista y es parte del coste que se ha de afrontar por la flexibilidad en la forma funcional de la función de producción. Conviene señalar, no obstante, que la sensibilidad de los outliers es claramente asimétrica: los resultados DEA son muy sensibles a las UTD que manifiestan muy buena práctica pero absolutamente insensibles a las que son manifiestamente ineficientes.
- **Soluciones:** 1) Asegurarse de que las UTD de la muestra son **homogéneas**, evitando comparar lo incomparable; si es preciso, imponer restricciones adicionales para garantizar que las comparaciones se limitan a UTD que ofertan servicios homogéneos y trabajan en un entorno similar. 2) Hacer diagnóstico de outliers y chequear la presencias de observaciones influyentes. En este sentido, trabajar con observaciones longitudinales de las mismas UTD puede ser muy ventajoso, así como hacer análisis de sensibilidad (calcular *qué pasaría si* eliminamos de la muestra determinadas UTD).

## Problema 3. Necesidad de controlar por la heterogeneidad de las UTD.

- Es éste el gran problema general de los modelos DEA. Puesto que los resultados pueden ser muy sensibles a la elección de UTD, hay que conseguir limitar las comparaciones a las UTD comparables, que operan en un entorno similar.
- **Soluciones:** Afortunadamente, los modelos DEA son lo suficientemente flexibles para solucionar fácilmente sus propias limitaciones. Una posibilidad empleada con relativa frecuencia consiste en introducir como **inputs nocontrolables**, con precio nulo, las variables de entorno que pudieran afectar al conjunto de posibilidades de producción de algunas UTD, dificultando o nublando las comparaciones.
- La **solución** general es **añadir restricciones** que limiten el conjunto de UTD estructuralmente comparables, que

producen outputs similares con inputs homogéneos y que operan en un entorno parecido (regulación, perfil mórbido de la población atendida).

#### **Problema 4. Cómo controlar por diferencias en calidad de la atención médica.**

- La mayor dificultad de los DEA es conseguir especificar los múltiples outputs que se dan en sanidad en términos de calidad constante.
- **Soluciones:** Muchas de las aplicaciones empíricas consideran la calidad. Los AED de atención primaria la incluyen de formas diversas, mediante la duración media de las visitas; el porcentaje de cumplimiento por los Centros de determinadas normas técnicas mínimas; madres satisfechas y muy satisfechas del servicio de atención perinatal y bebés de riesgo que sobreviven, etc.

#### **Problema 5. Definición adecuada de inputs y outputs.**

- Respecto a la definición de inputs y outputs, en hospitales se encuentra una gran variedad de opciones, que responden al debate conceptual sobre la medida del multiproducto hospitalario.
- El trinomio recursos (médicos, camas, quirófanos,...) / servicios (estancias, ingresos, intervenciones quirúrgicas,...) / salud suscita la tipología de aplicaciones siguiente: a) recursos producen servicios; b) servicios producen salud; y c) recursos producen salud.
- El primer tipo es con diferencia el más frecuente, aunque también el menos adecuado conceptualmente. Además, puesto que los mismos inputs y outputs sirven de base para el cálculo de medidas de productividad y de valores objetivo en hospitales públicos, el resultado del análisis podría estar sobreestimando sistemáticamente la eficiencia.
- Apenas hay en la literatura análisis del tipo c), que conceptualmente es más adecuado.

- Los inputs son financieros, recursos físicos o mixtos, dependiendo de los objetivos del análisis y de la disponibilidad de información. La introducción de recursos financieros lleva a estimar una mezcla de eficiencia técnica y asignativa, con lo que se pierde una de las ventajas potenciales de los modelos AED respecto a la estimación de fronteras estocásticas de costes. Algunas variables, como el número de MIR, actúan alternativamente como input o como output.
- **Soluciones:** El **análisis de sensibilidad** permite valorar la robustez de los resultados frente a cambios en la definición de inputs y de outputs.

## 5- Conclusiones

Las clasificaciones de centros, servicios y profesionales de la salud vienen siendo una práctica cada vez más habitual. La confección de una lista ordenada (ranking) de agentes proveedores de la salud tiene repercusiones financieras y de gestión en busca de la mejora de la efectividad y la eficiencia. También orienta a los pacientes para elegir centro.

Sin embargo, el uso de esa información debe realizarse con suma precaución. Las listas ordenadas de centros y servicios deben cumplir con un conjunto de requisitos relativos a la construcción de la escala de medida y a determinadas condiciones estadísticas y, en cualquier caso, acompañarse de información estadística que permitavalorar la significación de las diferencias.

Las principal ventaja de los métodos AED es su flexibilidad y adaptabilidad a una gran variedad de situaciones y contextos. La riqueza de resultados, con matices de considerable interés y la ausencia de hipótesis fuertes acerca del proceso de transformación de inputs en outputs son ventajas adicionales. Generalmente, se señalan como principales inconvenientes su carácter determinista y riesgo de confundir ruido aleatorio con ineficiencia, la sensibilidad de los resultados a la presencia de outliers en los datos y la imposibilidad de someter a contrastes de hipótesis estadísticos, rigurosos, los resultados.

Pero la ciencia todavía no garantiza que se obtengan medidas objetivas de eficiencia organizativa independientes de los instrumentos estadísticos. Hay demasiada sensibilidad de los ranking a pequeños cambios en el modelo o en los datos. Los resultados varían radicalmente, según se ajuste o no por factores que en principio se pueden considerar pertenecientes a un entorno fuera de control para los gestores. También suele ser difícil argumentar los resultados (justificar el ranking resultante con la lógica de la organización, y no con la lógica matemática). Significación estadística y significación socio-política ni son sinónimos ni siempre concuerdan.

La crítica más grave y certera se refiere a la elección y definición de inputs y de outputs y el ajuste por calidad, por severidad de los procesos atendidos y por servicios suplementarios de tipo hotelero. Es necesario refinar esas medidas, que muchas veces

son agregados demasiado burdos. Sin embargo, las posibilidades de desagregación están limitadas porque ha de haber un equilibrio entre el número de UTD y el de variables del modelo . Como el tamaño poblacional es finito y los hospitales sometidos a comparación han de ser homogéneos, la única solución es limitar razonablemente la desagregación de factores y de productos.

Si los efectos dinámicos son difíciles de novelizar en general, más todavía lo son en los modelos de eficiencia organizativa en sanidad. Los centros asistenciales arrastran la herencia del pasado —inercia no controlable— y toman decisiones hacia el futuro, gastando hoy para ganar mañana. La especificación dinámica de retardos y respuestas a impulsos es compleja, y requiere datos longitudinales que no siempre están disponibles o son comparables. El cambio tecnológico dificulta todavía más la comparabilidad temporal de la eficiencia organizativa.

En suma, los modelos que hemos presentado en este capítulo, y los modelos AED en particular, ofrecen un conjunto de instrumentos valiosísimos para comparar y medir la eficiencia productiva de Centros, Actividades y organizaciones sanitarias pero también encierran el peligro de que, si son aplicados incorrectamente, arrojarán resultados distorsionados.



## Referencias bibliográficas

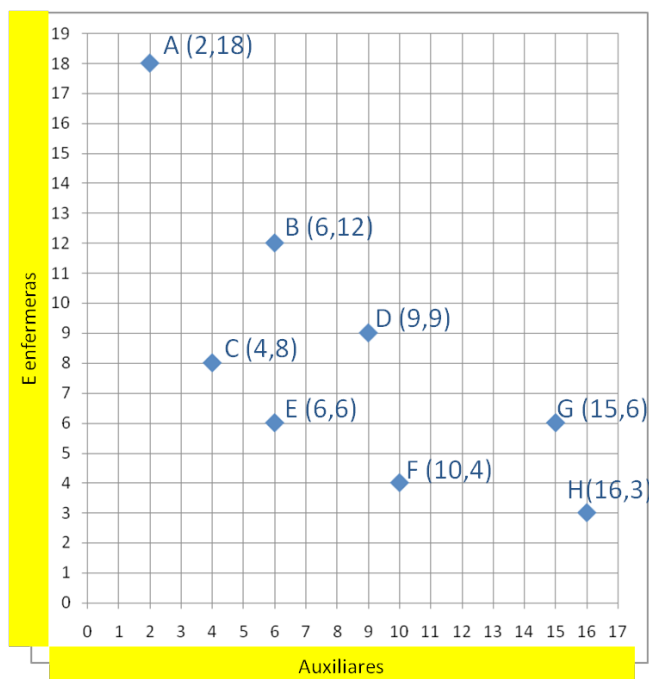
1. *Jaume Puig (2000) "EFICIENCIA EN LA ATENCIÓN PRIMARIA DE SALUD: UNA REVISIÓN CRÍTICA DE LAS MEDIDAS DE FRONTERA" Rev. Esp. Salud Pública v.74 n.5-6 Madrid set./dic. 2000*
2. *J Martin Martin y M. PUERTO LÓPEZ DEL AMO GONZÁLEZ "La medida de la eficiencia en las organizaciones sanitarias" Presupuesto y Gasto Público 49/2007: 139-161*

## ANEXO: un caso práctico simplificado.

Supongamos que tenemos 8 centros de vacunación; ajustado para un mismo nivel de producción (inmunizaciones), presentan combinaciones diferentes de "inputs", que se consignan como número de auxiliares y de enfermeras:

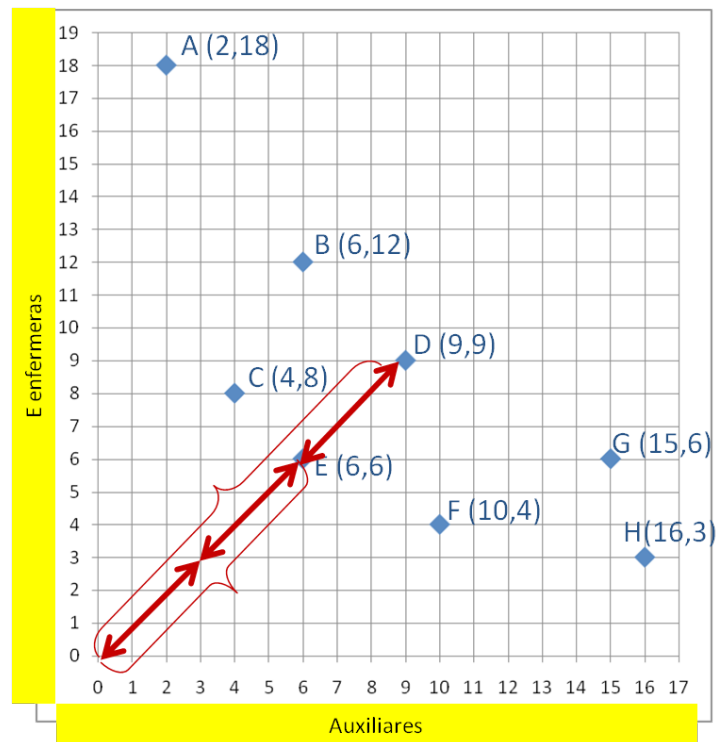
Centros	num. auxiliares	num. enfermeras
<b>A</b>	2	18
<b>B</b>	6	12
<b>C</b>	4	8
<b>D</b>	9	9
<b>E</b>	6	6
<b>F</b>	10	4
<b>G</b>	15	6
<b>H</b>	16	3

Si los representamos en dos ejes, con auxiliares den abscisas y enfermeras en ordenadas, vemos su distribución.

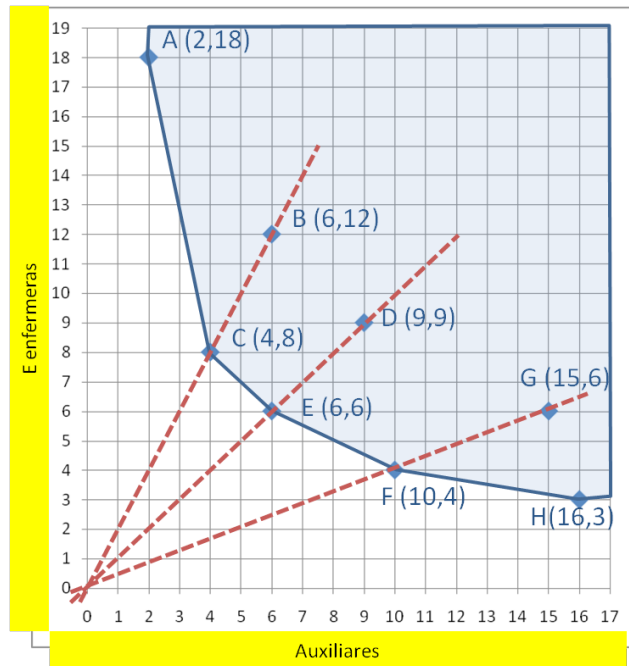


Si consideramos tan sólo los Centros E (6 auxiliares y 6

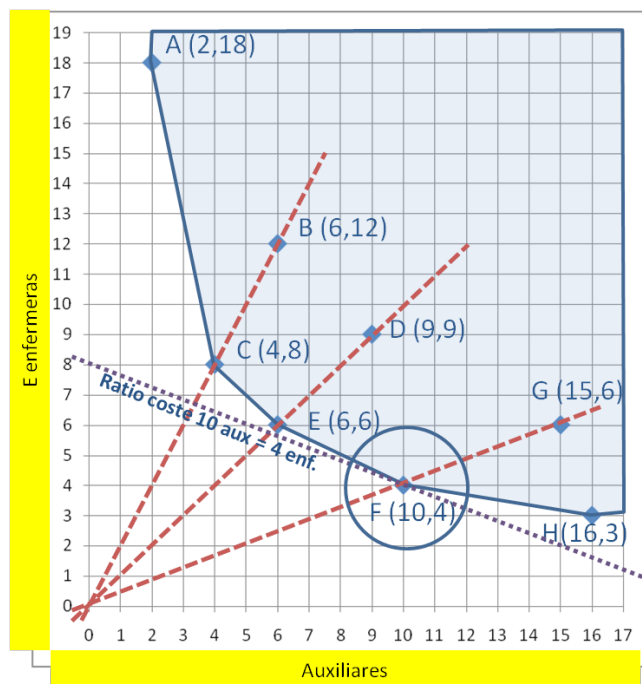
enfermeras) y D (9 auxiliares y 9 enfermeras), que tienen la misma proporción de ambos inputs, nos percatamos de que el D trabaja con menor eficiencia técnica, pues podría hacer lo mismo con menos factores de producción; al trazar un radio del origen de coordenadas O a los puntos E y D, vemos que la distancia OD excede en un tercio a la OE; esto expresaría gráficamente ese 1/3 de ineficiencia radial de Farrell.



Si trazamos una línea entre los Centros más eficientes, encontramos la frontera de eficiencia técnica; la zona oscura, fuera de la frontera, muestra los Centros cuya eficiencia es menor que otro centro que tenga la misma proporción de auxiliares y enfermeras (el B, D y G); los Centros A y H se consideran parte de la frontera, aunque no haya otros centros con su misma proporción). El análisis envolvente de datos, considera que cualquier centro con diferente proporción de personal, que esté dentro de la zona oscura, también sería ineficiente, y su distancia radial a un hipotético centro que estuviera en la frontera, determinaría su proporción de ineficiencia.



Finalmente, si contamos con el coste de auxiliares y enfermeras, podríamos decidir cuál Centro presenta un mejor uso del recurso económico (eficiencia asignativa en el uso de los factores de producción); por ejemplo, si suponemos que con el sueldo de una enfermera podemos contratar 2,5 auxiliares (o para mayor facilidad al dibujar, 4 enfermeras equivalen a 10 auxiliares), se trazaría una recta que representa dicha proporción; y donde esta *tangente* corta a la frontera de eficiencia, el Centro F, es donde se produce el mayor ahorro económico.



Lo que se demuestra haciendo estos cálculos:

Centros	num. auxiliares	num. enfermeras	coste aux	coste enf	coste total
			1	2,5	
A	2	18	2	45	47
B	6	12	6	30	36
C	4	8	4	20	24
D	9	9	9	22,5	31,5
E	6	6	6	15	21
F	10	4	10	10	20
G	15	6	15	15	30
H	16	3	16	7,5	23,5

## Epílogo del ejemplo:

Podemos tener un excelente resultado en el Centro F; pero si está rota la cadena de frío, resultará que se ponen las vacunas pero no se inmuniza; habría una ineficiencia asignativa en términos de resultados en salud (outcomes), por un fallo en la **efectividad**.

O incluso, podría haber una dudosa indicación de una determinada vacuna, cuya **eficacia** sea muy baja por sus propias características (cepa de gripe), o por desajuste con la prevalencia local (papiloma).