

Posibilidades y limitaciones de la gestión por resultados de salud, el pago por objetivos y el redireccionamiento de los incentivos. Informe SESPAS 2008

Salvador Peiró^a / Anna García-Altés^b

^aEscola Valenciana d'Estudis de la Salut, Valencia, España; ^bAgència de Salut Pública de Barcelona, Barcelona, España.

(Possibilities and limitations of results-based management, pay-for-performance and the redesign of incentives)

Resumen

El concepto de «pago por calidad» (P4P) reúne un conjunto de diferentes estrategias que intentan estimular la mejora de la calidad asistencial remunerando a los proveedores de atención sanitaria según sus resultados en el cumplimiento de objetivos de calidad o productividad predefinidos. Pese a que la efectividad del P4P para mejorar la calidad de la atención está poco establecida, estos sistemas se están difundiendo ampliamente en el Reino Unido, Estados Unidos y otros países, incluida España. Los elementos del diseño de los esquemas de P4P determinantes de su efectividad se refieren a quién debe recibir los incentivos, cuánto debe pagarse, qué debe retribuirse, la necesidad de incorporar ajustes de riesgos (sobre todo en los indicadores de resultados intermedios) y de tener en cuenta el clima organizativo, y la combinación óptima de incentivos financieros y no financieros. Las limitaciones más importantes que cabe considerar son: la orientación exclusiva hacia la reducción de la infrautilización, la afectación de la equidad, el efecto «lupa», la validez de los indicadores, la confusión entre recomendaciones de guías de práctica e indicadores de calidad, la ingeniería documental, el paternalismo con los pacientes, el posible impacto negativo sobre el profesionalismo y la motivación interna de los médicos, y el abordaje de la calidad como un problema de decisiones imperfectas, antes que como un sistema imperfecto. **Palabras clave:** Incentivos. Calidad asistencial. Indicadores de calidad.

Abstract

The concept of pay-for-performance (P4P) encompasses different strategies that aim to stimulate health care quality improvement by remunerating healthcare providers according to their performance in specific measures of efficiency or quality. Although the effectiveness of P4P in improving quality of care is largely unknown, these systems are being widely adopted in the United Kingdom, the United States and other countries, including Spain. The elements of P4P design that are most decisive for the effectiveness of these schemes are as follows: 1) *who* should receive the incentives, *how* they should be paid, *what* should be rewarded, the need to incorporate risk adjustments (mainly if surrogate outcomes are used as indicators) and the need to bear organizational climate and the optimal combination of financial and non-financial incentives in mind. The most important limitations to consider are the following: 1) the exclusive focus on reducing subutilization; 2) the effect on equity; 3) the “magnifying glass” effect; 4) the validity of indicators; 5) the confusion between the recommendations of clinical guidelines and quality indicators; 6) “document engineering”; 7) paternalism; 8) the negative impact on professionalism and clinicians’ internal motivation, and 9) the assumption that quality problems result from imperfect individual decisions rather than from an imperfect system.

Key words: Incentive. Quality of health care. Quality indicators.

Ideas clave

- Aunque la efectividad del «pago por calidad» (*pay for performance*, P4P) está poco establecida, este tipo de esquemas retributivos se están difundiendo rápi-

damente por todos los países desarrollados, incluida España.

- Quién debe recibir los incentivos, cuánto debe pagarse, qué debe retribuirse, y el papel de los incentivos no financieros, del ajuste de riesgos y del clima organizativo, son los elementos esenciales del diseño del P4P.
- Las limitaciones que cabe considerar incluye las siguientes: no considerar la sobreutilización, afectación de la equidad, efecto «lupa», escasa validez de mu-

Correspondencia: Salvador Peiró.
Escola Valenciana d'Estudis de la Salut.
Correo electrónico: peiro_bor@gva.es

chos indicadores, confusión entre recomendaciones de guías de práctica e indicadores de calidad, sobrerregistro e ingeniería documental, paternalismo, impacto negativo sobre la motivación interna, y el abordaje de la calidad como un problema individual, antes que organizacional.

- El Sistema Nacional de Salud necesita desarrollar nuevos indicadores capaces de orientar mejoras de salud y reducir desigualdades. Para ello, debe centrarse en cuidados de alto valor clínico y poblaciones de alto riesgo, y situar los incentivos de forma que favorezcan la coordinación de la atención.
- Reforzar la motivación intrínseca, el profesionalismo, el peso de los pacientes y la capacidad técnica de análisis y evaluación conforman el contexto esencial para el buen funcionamiento del P4P.

Introducción

En casi todos los países desarrollados hay una importante preocupación por la efectividad, la seguridad, la sensibilidad a las necesidades y las preferencias de los pacientes, y la eficiencia de la atención sanitaria. También hay el convencimiento de que los sistemas tradicionales de pago, tanto si se refieren a centros u organizaciones sanitarias, como si se refieren a médicos individuales, están ayudando muy poco a mejorar la calidad de la asistencia. A este respecto, un trabajo reciente señalaba:

«Existen muchos mecanismos para pagar a los médicos. Algunos son buenos y otros malos. Los tres peores son el pago por acto, la capitación y el salario. El pago por acto premia la provisión de servicios inadecuados, la sobrecodificación fraudulenta de visitas y procedimientos, y el “coleguismo” de las derivaciones en ping-pong entre especialistas. La capitación premia la denegación de servicios adecuados, el menosprecio de los enfermos crónicos, y una visión estrecha de la práctica médica que excluye a los pacientes cuya atención requiere mayor tiempo. El salario socava la productividad, fomenta las vacaciones en el trabajo y conduce a una mentalidad burocrática en la que cada procedimiento se convierte en un problema»¹.

Afortunadamente, es posible construir esquemas de retribución que combinen elementos de estos tres mecanismos e incorporen otros más sofisticados, para intentar alinear retribuciones y calidad asistencial. Estos diseños pueden incluir componentes prospectivos y retrospectivos, pueden poner diferente intensidad en distintos aspectos de la atención o adaptarse a las características de cada organización, pueden admitir incentivos no monetarios y tener en cuenta aspectos de desarrollo profesional más allá de los estrictamente retributivos y, también, pueden vincularse a indica-

dores concretos de calidad o eficiencia, sean de procesos o de resultados.

Los sistemas de «pago por calidad» o «pago por objetivos» (*value based purchasing, quality-based purchasing* o *pay for performance*, P4P) forman parte de estos nuevos sistemas de retribuciones a médicos (o a centros u organizaciones sanitarias) que intentan estimular la mejora de la calidad asistencial remunerando a los proveedores de atención sanitaria por alcanzar objetivos de calidad o productividad preestablecidos². Aunque los esquemas de P4P pueden ser muy variados, sus elementos comunes son: a) un conjunto de objetivos de calidad y/o eficiencia que delimitan el objeto evaluado; b) criterios y estándares explícitos para definir el cumplimiento de los objetivos, y c) incentivos (recompensas), típicamente monetarios, cuya cantidad global está en riesgo y depende del nivel de objetivos alcanzado por cada proveedor. Los incentivos pueden ofrecerse en forma de pagos extra por realizar determinadas actividades (una variante de pago por acto), gratificaciones globales por alcanzar determinados niveles de objetivos y, en ocasiones, pueden incluir penalizaciones por resultados negativos (p. ej., no reembolsar los costes adicionales derivados de complicaciones evitables).

La evidente racionalidad que subyace bajo la idea de que las retribuciones se usen para que los proveedores se esfuercen en mejorar la calidad de la atención ha convertido el P4P en un poderoso discurso. Apoyándose en la importante incorporación de tecnologías de la información al sector salud—sobre todo, en la historia clínica electrónica— los gobiernos y las aseguradoras de todos los países desarrollados se han apuntado a alguna estrategia de pago por calidad. En Estados Unidos, y aunque el P4P está muy lejos de ser una política generalizada³, la Veterans Administration y más de la mitad de las Health Maintenance Organizations (HMO) privadas disponen de algún programa de pago por objetivos, y el Congreso ha ordenado al Center for Medicare and Medicaid Services que desarrolle los trabajos necesarios para introducir este esquema de pago en Medicare⁴. En el Reino Unido se ha ido mucho más lejos. Desde 2004, el llamado New Contract entre el National Health Service (NHS) y los médicos generales vincula casi un tercio de las retribuciones totales de estos profesionales a sus puntuaciones en 146 indicadores de calidad y eficiencia⁵⁻⁷. En otros países también se están desarrollando algunas experiencias de P4P^{8,9}, aunque ninguna de ellas tiene la extensión de la británica.

En el Sistema Nacional de Salud (SNS) español, prácticamente todas las comunidades autónomas (CCAA) han desarrollado esquemas de «retribución variable» asociados a objetivos o indicadores de productividad y calidad que, en esencia, son formas más o menos sofisticadas de P4P. Estas retribuciones variables forman parte de los «contratos de gestión» y/o de los méritos que cabe considerar en la «carrera pro-

fesional» y, en general, son una evolución de los intentos iniciados en los años noventa por el INSALUD y algunas CCAA de incorporar «incentivos» económicos al rígido modelo salarial característico del SNS¹⁰. A diferencia del Reino Unido, donde hay un importante cuerpo de literatura médica (y un gran debate) sobre indicadores e incentivos, en el SNS apenas hay trabajos que discutan —en nuestro contexto— las ventajas y limitaciones teóricas de estos sistemas de pago¹⁰⁻¹⁵ y, mucho menos, trabajos empíricos que evalúen sus resultados^{16,17}. Esta situación es, por lo demás, habitual en el SNS, organización que combina una gran renuencia a la evaluación y limitaciones para el acceso a los indicadores (que son públicos o fácilmente accesibles en el Reino Unido y otros países) y, en consonancia con lo anterior, un mínimo desarrollo de la investigación en servicios de salud.

La escasez de marcos, evaluaciones o evidencias empíricas que soporten el P4P, o las limitaciones para construir indicadores específicos de calidad similares a los del NHS, no parecen haber entorpecido la apuesta política por la «retribución variable» en los importantes incrementos retributivos que en los últimos años se están produciendo en los servicios de salud de prácticamente todas las CCAA. Hay que señalar que en el SNS, los gestores y políticos (que perciben las retribuciones variables como una fuente de control sobre los colectivos profesionales), y los médicos (que las perciben como subidas salariales netas), han puesto mucho más empeño en la primera de las «P» (el «pago»), que en la segunda (la «calidad»). Incluso, y desarrollando una paradójica estrategia (*performance for pay* antes que *pay for performance*), en la mayor parte de las CCAA los incrementos han ido por delante de los indicadores, que tuvieron que ser construidos con premura para proceder a repartos retributivos previamente pactados. Fuera como fuera, a día de hoy todos los servicios de salud del SNS enfrentan (con tan pocas vacilaciones como información y análisis) los complejos problemas de cómo gestionar el pago por objetivos.

En este trabajo, se resumen las evidencias empíricas de mayor interés en nuestro contexto sobre la efectividad del P4P, se identifican los factores que deben tenerse en cuenta para valorar estas experiencias, y se describen las limitaciones más importantes de estos esquemas de pago. Finalmente, y como corolario de la exposición previa, se realizan algunas recomendaciones sobre el P4P en el SNS.

El pago por calidad, ¿mejora la calidad de la asistencia?

Contestar esta pregunta y, sobre todo, contestarla de forma útil para el SNS, no puede hacerse sin gran in-

certidumbre. Las revisiones sistemáticas disponibles¹⁸⁻²¹ son un claro ejemplo de las insuficiencias de esta metodología para dar soporte a la toma de decisiones de política y gestión sanitaria. La revisión de la Colaboración Cochrane sólo incluye 2 trabajos para la evaluación del P4P¹⁸. En el primero se incrementaba entre 0,80 y 1,20 dólares, sobre los 8 dólares previos, el pago por vacunación antigripal si se superaba un determinado porcentaje de cobertura vacunal, mientras que en el segundo se ofrecía un pago adicional a los médicos generales de una región escocesa que superasen un determinado porcentaje de cobertura en la vacunación pediátrica. En ambos casos mejoró el porcentaje de cobertura, aunque sólo en el primero el resultado fue estadísticamente significativo. Dudley et al^{19,22} encontraron 9 ensayos clínicos sobre P4P, aunque la mayoría no reportaba las características de los incentivos ni el contexto en que fueron aplicados. Las intervenciones evaluadas incluían el pago por acto de algunas actividades, los bonos o recompensas por objetivos de calidad y la información pública de datos comparativos sobre la calidad de los centros. De los 11 indicadores evaluados, 7 mostraron una respuesta positiva y significativa al P4P, pero no los otros 4.

La revisión de Petersen et al²⁰, abierta a trabajos sin aleatorización, incluía 17 estudios. Sus resultados muestran un poco de todo: bastante «ingeniería documental», algún caso de selección adversa de pacientes, mejoras reales pero que todavía se situaban lejos de los estándares y, también, algunos resultados más claramente positivos. La revisión de Rosenthal et al²¹, que incluye estudios del sector empresarial no sanitario, tampoco encontró evidencias empíricas sólidas que apoyaran la efectividad del P4P, aunque señalan que las intervenciones a pequeña escala, las más frecuentes en el entorno estadounidense, no parecen capturar la atención de los profesionales y son insuficientes para motivar cambios sustanciales en el comportamiento de médicos y hospitales. Por otro lado, estos autores sí encontraron evidencias de efectos no deseados, como el dolo, el trampeo (en programas de absentismo y en colegios), la selección de pacientes y la ingeniería documental. Petersen et al también remarcan la ausencia de estudios sobre otros posibles efectos negativos predichos por la teoría económica, especialmente sobre el «efecto lupa» y el descuido de las actividades no incluidas en los indicadores.

Las evaluaciones disponibles del New Contract del NHS británico^{6,7,23} apuntan una espectacular mejora en los indicadores evaluados, hasta el punto que el gobierno británico fue sorprendido por el nivel de objetivos alcanzado y tuvo que aportar mucho más dinero que el inicialmente previsto²⁴. Más dudosa es la atribución del conjunto de las mejoras al New Contract, ya que muchos indicadores seguían tendencias de mejora establecidas desde años antes y, como mucho, el P4P sería respon-

sable de una discreta aceleración de estas mejoras⁷. Esta posibilidad viene reforzada porque algunos indicadores no incluidos en el P4P mejoraron de forma similar a los incluidos⁷. Adicionalmente, hay evidencias de «ingeniería documental» en los «incrementos de calidad»^{23,25}, tanto en lo que se refiere a la «mejora» de los registros con importancia para el cumplimiento de objetivos, como por la exclusión del denominador de los indicadores de pacientes previsiblemente no cumplidores.

Respecto a los dos estudios empíricos en el SNS, ambos en Cataluña, en el primero los centros participantes mostraron niveles de calidad excelentes (aunque es poco valorable, ya que se trataba de centros de referencia para construir estándares de calidad)¹⁷, mientras que el segundo encontró discretas mejoras en el clima laboral, pero sin cambios en la satisfacción de los pacientes¹⁶. También hay informaciones anecdóticas sobre artimañas en los registros y, aún más importante, sobre la complicidad de los directivos y de prácticamente toda la organización sanitaria en una especie de «conspiración por la excelencia» que se articula en torno al sobrerregistro y el registro sesgado de las actividades monitorizadas por los indicadores, lo que conlleva que «[...] las cifras de control de la presión arterial de todos los equipos de atención primaria de Barcelona incluidos un proceso de *benchmarking* sean superiores a los de los centros excelentes adscritos voluntariamente al Programa de Actividades Preventivas y de Promoción de la Salud (PAPPS) de la Sociedad Española de Medicina de Familia y Comunitaria, o a los de los países de nuestro entorno»²⁶.

En conjunto, la literatura médica que ha revisado las evidencias empíricas sobre la efectividad del P4P para mejorar la calidad de la atención sanitaria puede resumirse de la forma siguiente:

- Trabajos, en general de baja calidad metodológica, con resultados inconsistentes y contradictorios respecto al impacto del P4P en la mejora de calidad. Sobre su coste-efectividad, simplemente, no hay estudios.

- Centrados en sólo dos países, Estados Unidos y Reino Unido, con sistemas de atención, esquemas retributivos básicos y modelos de P4P muy diferentes. Como norma, los trabajos estadounidenses evalúan intervenciones de P4P muy específicas y casi siempre preventivas, como los sobrepagos por vacunar, el cribado de cáncer de cérvix o la revisión del fondo de ojo en pacientes diabéticos. Estas intervenciones fueron realizadas en su mayor parte sobre poblaciones u organizaciones concretas (en algunos casos incluso sobre médicos residentes) y sus evaluaciones, típicamente, están referidas a un solo indicador de calidad relacionado con el motivo de pago (como la cobertura de vacunación antigripal), o a un solo aspecto de la atención (como la diabetes del adulto). La experiencia del NHS, sin embargo, referida a una intervención generalizada sobre toda

la atención primaria y con un amplio panel de indicadores, es de mayor interés para el entorno español.

- Hay evidencias, más o menos anecdóticas, que sugieren el desarrollo de los efectos adversos previsibles desde la teoría (sobrerregistro, selección de riesgos, y otros), aunque no se conoce la importancia global de estos problemas y no es posible realizar un balance entre estos efectos adversos y los efectos positivos.

- En prácticamente todos los trabajos de investigación evaluativa, los agentes que han puesto en marcha el programa de P4P han escogido la modalidad de incentivos y los objetivos de calidad que creían más adecuados para la organización sobre la que actuaban y en el contexto concreto en que lo hacían, y no es obvio que las mismas intervenciones tengan similar impacto en otros entornos.

En conjunto, estas características limitan enormemente la validez externa de las evaluaciones disponibles en la literatura médica que, aun siendo reveladoras, resultan una guía muy endeble para generalizar las estrategias de P4P a otros contextos. Esta conclusión, que es probablemente aplicable a los propios países en los que se desarrollaron los estudios, es incuestionable en entornos tan idiosincrásicos como el español, con una colectivización de las negociaciones salariales, una cultura de falsa equidad y «café para todos», un escaso peso del profesionalismo, un escaso peso –o contrapeso– de los pacientes, una elevada politización –que alcanza a los niveles más rasos de la cadena de gestión– pero sin obligaciones de transparencia ni rendición de cuentas, CCAA que tienden a emular a las que más pagan, base de pago por salario tanto en atención especializada como en atención primaria, y demás rasgos que hacen del SNS un sistema muy peculiar y diferente al de otros países, incluido el NHS británico.

Elementos del diseño en el pago por objetivos

Las dificultades para la generalización de los resultados de los estudios comentados no es óbice para que estas experiencias no tengan un alto interés para el SNS. Desde la elaboración y la selección de indicadores, a la valoración de las condiciones que se asocian al éxito o el fracaso de estas intervenciones, el análisis de su desarrollo es altamente informativo. A la hora de valorar las posibilidades de generalización de los estudios a los diferentes entornos y, también, para valorar las posibilidades de desarrollo de esquemas de P4P propios, conviene extraer de esta literatura médica los elementos del diseño que serán determinantes de su efectividad y, también, de sus efectos adversos. Las preguntas relevantes incluyen el quién, el cuánto y cómo, el qué y algunas otras^{9,21,27-31}.

¿Quién debe recibir los incentivos?

Una primera aproximación a esta pregunta se refiere a si los incentivos deberían enfocarse hacia grupos de profesionales conjuntamente, hacia médicos individuales, o ambos. Los incentivos de grupo distribuyen los riesgos entre los profesionales que trabajan en un servicio, en un centro o en un equipo de atención primaria y fomentan el control por «pares», pero son más débiles que los individuales que, por otro lado, presentan problemas estadísticos, dado el bajo tamaño muestral, y pueden estar muy afectados por el error aleatorio. La combinación de ambas perspectivas puede tener interés, especialmente cuando se trabaja en equipo.

Una segunda aproximación a esta pregunta parte de que los resultados sanitarios suelen implicar la contribución de diversos profesionales. Los resultados obtenidos por los diabéticos, por ejemplo, no sólo dependen de la atención que prestan los médicos generales, en los que se centran indicadores e incentivos. Médicos especialistas (en endocrinología, oftalmología, cardiología, cirugía vascular periférica, dermatología, etc.), enfermeras y otros profesionales son importantes en muchos indicadores. La ubicación de indicadores e incentivos en un responsable asistencial inadecuado (o parcial) puede reducir notablemente el impacto del P4P sobre los resultados clínicos. En el entorno español es muy probable que muchos de los incentivos situados en atención primaria (sobre todo los de prescripción) deban situarse –al menos, también– en la atención especializada, con la idea esencial de orientarlos para favorecer la atención coordinada e integral de los pacientes.

¿Cuánto es suficiente?

En términos de posología (cantidad, frecuencia, duración) la relación dosis-respuesta de los incentivos a médicos es desconocida, aunque se sabe que tiene relación con las retribuciones totales y con las posibles alternativas de obtención de otros ingresos. Los incentivos, en teoría, para ser efectivos, deben compensar el esfuerzo realizado y deben suponer una parte sustantiva del conjunto de ingresos. En muchos programas de P4P, sin embargo, las cantidades que recompensan la calidad son mínimas o suponen una parte muy pequeña de los ingresos de un médico, con lo que su influencia en los resultados es, previsiblemente, baja. Este aspecto es una limitación muy característica de los estudios estadounidenses en los que un médico (que atiende usuarios de diversas aseguradoras) puede tener bajo esquemas de P4P a porcentajes muy pequeños de pacientes (p. ej., el 5-10 de sus diabéticos y menos del 1% de sus visitas), con un impacto esperable, incluso entre esos pacientes, muy pequeño. Las posibilidades de obtener ingresos de fuentes

alternativas (p. ej., dedicando mayores esfuerzos a la atención de pacientes privados) también influyen en el impacto de los incentivos, que tienen que competir con estas otras fuentes de ingresos por el tiempo y el esfuerzo de los médicos. Del mismo modo, los incentivos tendrán más efecto cuando se implanten sobre esquemas básicos de pago capitativo o por salario (como en el SNS) que cuando se implanten sobre esquemas básicos de pago por acto (como las aseguradoras privadas).

¿Qué retribuir?

El P4P requiere la evaluación cuantitativa del trabajo de los clínicos, usualmente en función del porcentaje de pacientes de un subgrupo concreto (denominador) que experimenta un determinado resultado (numerador), o de la adhesión del médico a determinados estándares de proceso que se pretenden «basados en la evidencia». El mantra de las medidas de calidad dice que, idealmente, estas deben ser clínicamente relevantes, válidas (racionalidad clínica), estables al menos durante unos años (para permitir comparaciones a lo largo del tiempo), factibles, sólidas (fiables) y capaces de orientar la organización hacia la mejora³².

Los problemas para desarrollar este tipo de mediciones del trabajo clínico son muy importantes: a) ausencia de criterios de evaluación basados en la evidencia para muchos tipos de cuidados y para muchos tipos de pacientes; b) incertidumbre en la definición de los límites de la enfermedad y entre cuidados aceptables e inaceptables; c) escaso tamaño muestral y otros problemas estadísticos, que se incrementan enormemente cuando se quiere valorar a médicos individuales o la atención de subconjuntos pequeños de pacientes; d) diferencias en la gravedad de los pacientes atendidos por cada profesional o cada centro, que puede determinar notables diferencias en la mayor parte de las medidas de proceso o de resultado; e) escasez de evidencias sobre cómo serían los procesos óptimos de atención, especialmente el manejo longitudinal a largo plazo de pacientes, y f) factibilidad y costes, incluido el tiempo que los médicos y el personal de enfermería deben dedicar a registrar datos.

Adicionalmente, la fijación del criterio o de los puntos de corte de cada indicador que sirven para definir la calidad o no de la asistencia prestada es también un aspecto complejo, sobre todo cuando se transforman medidas derivadas de escalas continuas (presión arterial, colesterolemia, hemoglobina glucosilada [Hb1_{Ac}]) en dicotómicas (p. ej., porcentaje de pacientes con una presión arterial inferior a 130-140 mmHg, Hb1_{Ac} menor de 7, 8 o 9). El establecimiento de criterios estrechos puede implicar la hipermedicalización de los pacientes, y también la desmoralización de los médicos que partían de posiciones más lejanas a los óptimos. Unos criterios de-

masiado laxos suponen que muchos centros, con espacio potencial de mejora, habrían alcanzado de antemano sus objetivos finales. Una variante de este problema se plantea cuando hay que optar entre retribuir por alcanzar un determinado nivel de calidad (p. ej., que el 80% de los diabéticos tipo 2 tenga valores de Hb1_{Ac} < 9) o por mejorar (haber pasado del 40 al 70% de pacientes en el indicador previo, pero sin alcanzar el estándar óptimo). En este caso, un desequilibrio entre componentes puede «desmotivar» a los que alcanzan los óptimos (si sus compañeros con peores niveles de calidad reciben más incentivos) o las mejoras parciales de los médicos o equipos con peores indicadores.

Cuántas medidas incluir y cuáles son las fuentes de datos, su solidez, fiabilidad y facilidad de manipulación, son cuestiones añadidas a la pregunta de qué retribuir. Para tener éxito, un sistema de P4P debe conseguir un buen equilibrio en «el qué», de forma que contenga la manipulación y no desmoralice a los «mejores» profesionales, pero tampoco desincentive la mejoría en los «peores».

¿Se requieren ajustes por gravedad?

Los pacientes con niveles educativos altos, buen nivel económico y apoyo familiar suelen cumplir mejor los tratamientos, tener menos fallos de seguimiento (muchos indicadores derivan de acudir a revisiones, como realizar un fondo de ojo anual en los diabéticos) y obtener mejores resultados que los pacientes de características contrarias. Por otra parte, mientras requiere poco esfuerzo conseguir que un adulto joven disminuya su presión arterial de 141 a 139 mmHg y conseguir el objetivo del P4P (basta con tratarlo con cualquier antihipertensivo, incluso aunque sea un tratamiento dudoso) es mucho más complejo conseguir que un paciente de 80 años con hipertensión sistólica del anciano consiga cifras < 140 mmHg (pese a emplear combinaciones de 3-4 antihipertensivos, y otros tratamientos para sus efectos secundarios). Si los indicadores no están contruidos sobre subgrupos de pacientes homogéneos (o, como en el caso británico, se permite la exclusión de los «malos riesgos»), el P4P puede perjudicar a los médicos que traten «peores» pacientes (más ancianos, con peor nivel socioeconómico, sin recursos familiares, etc.), favorecer la exclusión del sistema de estos grupos poblacionales y reducir la calidad de su atención.

Clima y cultura profesional y organizacional

Las características y el clima de cada organización —en realidad, incluso de cada microsistema de la misma organización— pueden influir notablemente en el impacto del pago por objetivos. Los microsistemas incluyen las relaciones con los pacientes y una imagen inadecua-

da de algunos de estos incentivos (especialmente, los pagos vinculados a las menores tasas de derivaciones o de prescripción) entre los pacientes, también podría afectar a los resultados del P4P.

¿Cuál es la combinación óptima de incentivos financieros y no financieros?

Las estrategias no retributivas para la mejora de la calidad de la atención (guías, *audit*, *feedback*, procesos, difusión pública de los resultados) también han demostrado contribuir a la mejora de los indicadores, y tienen un importante potencial en su empleo aislado o combinado con el P4P, tal vez con efectos sinérgicos. Además, los otros «incentivos» (orgullo profesional, satisfacción por el trabajo bien hecho, reconocimiento, oportunidades —sobre todo en atención especializada— para trabajar en áreas o unidades «punteras» o, simplemente, ampliar las posibilidades de formación) pueden tener tanto o más peso que los económicos e incluso ser determinantes para crear un clima en que los incentivos monetarios puedan actuar positivamente.

Resumiendo, los aspectos anteriores, aunque el análisis de los estudios —aleatorizados u observacionales— tiene valor para intentar identificar situaciones asociadas al éxito o fracaso del P4P, tienen una escasa validez externa y las posibilidades de generalizar sus resultados (incluso los del NHS) al entorno español son muy dudosas. Un análisis del diseño del P4P que se pregunte por sus elementos clave (quién, cuánto, cómo, qué, dónde, junto a qué otras cosas) puede ser mucho más informativo para considerar los posibles efectos del P4P en el contexto del SNS y sus diferentes entornos (atención primaria y especializada, hospitales comarcales y terciarios, enfermedades crónicas y agudas, consulta y domicilio, urgencias y prevención primaria). Una evaluación desapasionada de las experiencias en el SNS, aun siendo poco esperable en el panorama político habitual, sería muy instructiva para la toma de decisiones.

Limitaciones de la gestión por «indicadores» y el pago por objetivos

El P4P tiene limitaciones, y también puede tener consecuencias no deseables que conviene tener en cuenta. Las más relevantes serían las que citamos a continuación.

Orientación a la reducción de la infrutilización

La mayor parte de las iniciativas de P4P se han centrado en incrementar la utilización de tratamientos efectivos (p. ej., bloqueadores beta en prevención secun-

daría del infarto de miocardio) o de medidas preventivas basadas en el consenso profesional (realización de una ecocardiografía anual en la insuficiencia cardiaca), pero prácticamente no hay iniciativas para reducir la sobreutilización. Esta orientación del P4P hacia la resolución de problemas de subutilización podría agravar los problemas de sobreutilización. Así, se ha señalado que los indicadores de «calidad», como el de «administración de antibióticos en las primeras 4 h de la llegada al hospital, en la sospecha de neumonía», pueden llevar a que reciban antibióticos todos los pacientes con fiebre y tos en la puerta de urgencias³³.

Recientemente se ha recomendado expandir el enfoque P4P para intentar mejorar la adecuación de la cirugía discrecional «sensible a las preferencias de los pacientes»³⁴ y reducir las hospitalizaciones por enfermedades crónicas (un tipo de atención muy «sensible a la oferta de recursos hospitalarios»)³⁵, pero no parece haber avances prácticos en este terreno. De hecho, y dado que políticamente no parece muy inteligente retribuir a los médicos —al menos a los especialistas— por hacer menos de algo (incluso cuando hacer menos signifique prestar atención de mejor calidad), es previsible que el problema de la sobreutilización de servicios no efectivos (o incluso perjudiciales) deba abordarse a través de otras aproximaciones de «retribución por calidad»³⁶.

Afectación de la equidad

La equidad es una dimensión importante de la calidad de los servicios sanitarios y un atributo clave de su *performance*. Para orientar las organizaciones hacia la reducción de las desigualdades, los indicadores deberían estratificarse por nivel socioeconómico, sexo o grupo étnico, un aspecto generalmente olvidado en el P4P³⁷. Los estudios que han revisado estos aspectos, tanto en Estados Unidos como en el Reino Unido, han mostrado que las consultas de las áreas más desfavorecidas muestran peores indicadores³⁸⁻⁴¹, aspecto que penalizaría a los centros que atienden estas áreas⁴² y, si se vinculan retribuciones a estos indicadores, generaría incentivos para intentar la exclusión de estos pacientes. No obstante, los estudios británicos tienden a mostrar algunas mejoras de los indicadores en los grupos más desfavorecidos (aunque menores que las mejoras en los niveles sociales más altos), lo que sugiere que los esquemas de P4P podrían mejorar la atención de los más desfavorecidos, aun a costa de un incremento de las desigualdades³⁸.

«Efecto lupa»

Los indicadores que soportan los sistemas de P4P se concentran, casi necesariamente, sobre algunas actividades clínicas sencillas de medir. Estas actividades

incluyen la prescripción de determinados medicamentos (en buena medida, para la prevención primaria o secundaria) y en la realización de actividades preventivas, sobre todo en algunas poblaciones de pacientes con enfermedades crónicas⁴³. Un resultado indeseable de esta situación es la sobrededicación a servicios —en muchas ocasiones, de reducido valor clínico—, cuya visión es ampliada por el «efecto lupa» de los indicadores, con descuido de otros aspectos igualmente importantes (y, en muchos casos, más importantes)^{14,44}. En este sentido, algunos trabajos apuntan que:

– La prestación de los servicios preventivos recomendados como prioritarios por las guías de práctica clínica supondría, para un médico de familia típico, un total de 7,4 h de trabajo por día laborable⁴⁵. Igualmente, la atención de las 10 enfermedades crónicas más prevalentes siguiendo las guías usuales, supondría para estos médicos 3,5 h de trabajo diarias⁴⁶. Se trata de un aspecto de especial interés porque los sistemas de indicadores utilizados en el P4P descansan precisamente sobre estos tipos de actividades, cuya realización no dejaría tiempo para ninguna otra actividad (en especial para atender a los enfermos agudos o las descompensaciones agudas de los pacientes crónicos).

– Tres cuartas partes de la población británica no padece ninguna de las enfermedades incluidas en el NHS Quality and Outcomes Framework⁴⁷, pero desde su introducción los grupos con intereses especiales han intentado incluir sus particulares prioridades entre los objetivos medidos y retribuidos. Por el contrario, no se consideran otros problemas prevalentes y potencialmente graves (p. ej., la depresión o la esquizofrenia)^{48,49}.

– Buena parte de los indicadores de calidad utilizados, incluso por el NHS⁵⁰, se refieren a la prestación de cuidados de valor no demostrado, y este tipo de atención compite por los recursos disponibles con otros tipos de atención más prioritarios.

Las preguntas obvias son: ¿qué pasa con lo que no se mide?, ¿cuáles son los costes de oportunidad de la sobreatención a las actuaciones médicas que disponen de indicador? Algunos autores que han utilizado indicadores «ocultos» o «no retribuidos» para valorar el «efecto lupa» no encontraron evidencia de empeoramiento de la atención no sujeta a P4P^{51,52}, aunque se trata de estudios muy específicos y de compleja interpretación (p. ej., en uno de ellos tampoco se encontró evidencia de mejora en las medidas sujetas a P4P), por lo que sus resultados no deberían generalizarse.

Validez de los indicadores

El valor y los costes de oportunidad de dedicarse a tratar los problemas que se han incluido en los sis-

temas de P4P dependen, sobre todo, de la validez de los indicadores utilizados. El concepto de validez, en este contexto, se refiere a la existencia de una relación causal sólida entre los niveles alcanzados en los indicadores utilizados para medir-retribuir la «calidad» (usualmente medidas de proceso y resultados fisiopatológicos) y mejoras reales en resultados clínicos relevantes (como la mortalidad, la incidencia de episodios mórbidos, la reducción de complicaciones relevantes, las mejoras funcionales o de calidad de vida y otras medidas similares). La literatura médica a este respecto es, sin embargo, poco consistente:

– En una evaluación del programa de «demostración» de P4P más importante de Medicare-Medicaid se analizaron los resultados de más de 100.000 pacientes con infarto agudo de miocardio (IAM) en 54 hospitales bajo P4P, comparándola con un grupo control de más de 400 hospitales sin pago por calidad. Los pacientes bajo P4P mostraban un porcentaje discretamente mayor de prescripción de aspirina al alta y de consejo de abandono tabáquico, pero no había diferencias en el conjunto de medidas de proceso para el IAM del American College of Cardiology/ American Hospital Association (AHA/ACC), ni en la mortalidad intrahospitalaria⁵².

– Ningún indicador de calidad en el manejo hospitalario de la insuficiencia cardíaca congestiva (ICC) del AHA/ACC se asoció a reducciones de mortalidad a los 60 y 90 días del ingreso hospitalario, aunque uno de los indicadores (la administración de inhibidores de la enzima conversiva de la angiotensina [IECA]), sí se asoció a la mejora de un resultado combinado de mortalidad y/o reingreso. Por otra parte, la prescripción de bloqueadores beta al alta (que no era una medida de calidad de las organizaciones mencionadas) se asoció con una importante mejora de estos resultados⁵³.

– En otro trabajo estadounidense, realizado sobre los más de 3.500 hospitales de agudos participantes en un programa para dar a conocer sus indicadores a los pacientes, las instituciones que se situaban en el cuartil más elevado en indicadores de calidad para el IAM, la ICC y la neumonía, mostraban, respecto a los tres cuartiles con peores indicadores, reducciones absolutas de mortalidad entre 0,001 y 0,012 según el proceso y la medida de mortalidad (intrahospitalaria, a los 30 días y al año del ingreso). Las comparaciones en el otro sentido (tres cuartiles superiores frente al peor cuartil) mostraban cifras de reducción absoluta de los indicadores de mortalidad de 0,008 y 0,018 para el IAM, incrementos del 0,013 y 0,038 en la mortalidad por ICC y reducciones de 0,003 y 0,014 para la neumonía⁵⁴. Aunque dado el elevado número de hospitales del trabajo algunas de estas cifras fueron estadísticamente significativas, su relevancia clínica es mínima⁵⁵.

– Wang et al⁵⁶ compararon, en pacientes con IAM, la mortalidad estandarizada de los 50 «mejores» hos-

pitales de un *ranking* estadounidense con el resto de centros. Además, los «mejores» hospitales tenían un menor riesgo ajustado de muerte, pero las tasas de mortalidad de estos centros se superponían con las del resto, y unos pocos de los «mejores» tenían tasas ajustadas de mortalidad de entre las peores de Estados Unidos.

– Bradley et al⁵⁷, en casi 1.000 hospitales que participaban en un registro de IAM, encontraron asociaciones moderadas o débiles entre diversas medidas de proceso (aspirina al ingreso, bloqueadores beta al ingreso, al alta, IECA al alta, consejo antitabaco, tiempo hasta la reperfusión). Algunas de estas medidas se correlacionaron con las tasas de mortalidad a los 30 días del IAM, pero apenas alcanzaron a explicar un 6% de la varianza en esta medida de resultado.

– En la experiencia británica, Bottle et al⁵⁸ encontraron relaciones muy débiles entre las puntuaciones de calidad de 303 equipos de atención primaria y las tasas de hospitalizaciones por enfermedad coronaria. Un trabajo similar encontró asociaciones pequeñas e inconsistentes (ocasionalmente, mejor «calidad» se asociaba a peores resultados clínicos) entre las puntuaciones de calidad y los ingresos por urgencias o la mortalidad⁵⁹.

Aunque también hay estudios en que los indicadores de calidad sí se asociaron a mejores resultados clínicos⁶⁰, la ausencia de una clara relación entre ambos tipos de medidas es patente; y ello pese a las muestras con gran número de hospitales, el empleo de sofisticados sistemas de ajuste de gravedad y el análisis con complejos métodos estadísticos. Esta situación plantea numerosas incógnitas sobre el uso de estas medidas de calidad en los esquemas de P4P. No tanto porque los proveedores no deban esforzarse en emplear determinados tratamientos (que, en buena parte aunque no todos, habían mostrado una sólida relación con los resultados clínicos en los homogéneos grupos de pacientes de los ensayos clínicos controlados)⁵⁵ sino porque los beneficios esperables en la práctica real (con grupos de pacientes heterogéneos, y respecto al manejo usual empleado por cada proveedor adaptándose en el seguimiento a cada paciente) serán menores que los relatados en los ensayos y, en muchas ocasiones, mínimos.

Confusión entre las recomendaciones de guías de práctica e indicadores de calidad

En los esquemas de P4P buena parte de los incentivos se asignan en función de la adhesión de médicos o centros sanitarios a determinadas medidas que, en muchas ocasiones, se han derivado de recomendaciones de guías de práctica clínica (GPC) basadas en estudios clínicos con aleatorización. Estas medidas

basadas en evidencias sólidas tienen más credibilidad clínica y suscitan mayor confianza que otros indicadores. Lamentablemente, la conversión de las recomendaciones de las GPC en indicadores de calidad es muy problemática y puede tener importantes efectos adversos^{61,62}.

Las GPC son conjuntos de recomendaciones para aplicarse prudentemente según la experiencia clínica y cuyo objetivo básico es facilitar una síntesis de información para reducir la brecha entre conocimiento científico y práctica clínica. Para ser comprensivas de diversas situaciones, mantienen un lenguaje flexible, reconociendo zonas grises de elevada incertidumbre, la complejidad de la atención médica y la necesidad de incorporar las preferencias de los pacientes. Aconsejan, instruyen y sugieren cuidados y actuaciones, sin forzar su cumplimiento. Los indicadores de calidad son criterios cuantitativos que indican el rendimiento en el cumplimiento de un determinado objetivo y cuya intención es evaluar la calidad de la atención prestada por médicos o instituciones. Su lenguaje es rígido, con criterios específicos de lo que es correcto y lo que es incorrecto, usados en forma de algoritmos muy simples a partir de datos fácilmente disponibles en la mayor parte de las organizaciones. Y tienen carácter forzoso, asociado a incentivos o penalizaciones.

Por ejemplo, la recomendación de intentar que la presión arterial de los diabéticos se sitúe por debajo de 130/80 mmHg es usual en las GPC; pero por sus propias características (el concepto de «guía», de recomendación no forzosa), las GPC admiten que cambiar a tratamientos más intensivos para conseguir este objetivo depende de diversos factores, incluidas las circunstancias del paciente (riesgo basal, comorbilidad, tratamientos concomitantes, carga de los nuevos tratamientos) y los propios riesgos de los nuevos tratamientos. Por ejemplo, dos diabéticos con idénticas cifras de hipertensión moderada (supongamos 140/90 mmHg) podrían tener un riesgo basal de muerte cardiovascular del 20% a los 5 años (varón de 65 años de edad, fumador, diagnosticado hace 15 años, con una Hb1_{Ac} del 8% y un colesterol total de 200 mg/dl) o del 1% (mujer de 50 años de edad, no fumadora, con una Hb1_{Ac} del 7% y un colesterol de 180 mg/dl). Un tratamiento más intensivo que redujera en un 10% el riesgo de muerte cardiovascular sería importante para el primer caso, pero no tanto para el segundo⁶³. Estas situaciones y otras similares hacen que el (buen) juicio clínico permita individualizar los tratamientos y obviar las recomendaciones genéricas de las GPC, sin que ello pueda valorarse como atención de baja calidad⁶⁴ o calificarse de «inercia clínica»⁶⁵.

Los indicadores de calidad, por el contrario, deben estar conformados, sobre todo, por prioridades de asistencia sanitaria para mejorar la atención a un determinado problema. Estos indicadores, simples y fáci-

les de obtener, no pueden incluir a todos los diferentes tipos de pacientes que se agrupan bajo diagnósticos como diabetes o ICC, ni intentar evaluar todos los aspectos de la atención a un determinado problema, pero sí pueden y deben orientarse hacia los pacientes y procesos que se presumen más sensibles a la mejora de los resultados clínicos. Por ejemplo, incluso reconociendo que una Hb1_{Ac} < 7% fuera un objetivo terapéutico correcto, también es importante conceder que unos valores del 7-8% son un problema mucho menor que unas cifras del 10%. Un indicador que concentre la atención en pacientes con el primer grupo de valores puede producir menos beneficios que otro que se centre en pacientes con mayor riesgo. En la práctica, las pequeñas desviaciones sobre el objetivo «teórico» pueden carecer de importancia mientras que los efectos adversos para lograrlos pueden ser importantes. Por ejemplo: utilizar 2-3 antidiabéticos orales, 2-4 antihipertensivos y dosis altas de estatinas para que un diabético de 80 años con hipertensión e hiperlipemia alcance los puntos de corte en los indicadores de diabetes puede no ser una estrategia clínica adecuada.

La transformación de las recomendaciones de la GPC en indicadores rígidos (tipo «todo o nada») tiende a priorizar la prestación de cuidados de mínimo valor en pacientes de bajo riesgo y, por ello, no deben esperarse grandes beneficios de esta estrategia. Por el contrario, este tipo de indicadores premian el uso de tratamientos especulativos⁶⁶, caros para el sistema sanitario, incómodos para los pacientes y de desconocida seguridad a largo plazo^{44,67-69}. Esto mismo sucede con muchas de las prácticas preventivas recomendadas en las GPC y transformadas en indicadores, como el examen anual del fondo de ojo en pacientes diabéticos cuyo examen previo fue normal^{70,71} o la ecocardiografía anual en pacientes con ICC estables y sin cambios clínicos²⁵. Las recientes polémicas en Estados Unidos para fijar un nuevo punto de corte en el indicador de Hb1_{Ac} del National Committee for Quality Assurance (NCQA) han sacado a la luz el componente «político» de estos indicadores, al poner de manifiesto el interés de la industria farmacéutica (que financió una campaña para modificar el indicador previo)⁶⁸, pero también de las asociaciones de enfermos y de las sociedades de especialistas (que buscan atraer recursos hacia «su» enfermedad o «su» especialidad) en confundir recomendaciones e indicadores.

Ingeniería documental y marrullerías con los datos

Como señalaba un médico británico, «la única forma de beneficiarse de los incentivos financieros contenidos en el nuevo General Medical Services Contract es optimizar el registro de datos de las perso-

nas incluidas en los objetivos del Department of Health»⁷². «Optimizar» no quiere decir, al menos necesariamente, incrementar la exhaustividad y la calidad de las historias clínicas, sino seleccionar los datos (y pacientes) que maximizan el reembolso. En la experiencia británica diversos datos apuntan que tanto el registro selectivo de los datos favorables como y, sobre todo, la exclusión de los denominadores de los pacientes que no alcanzaban los objetivos han sido elementos importantes en la «mejora» experimentada por los indicadores del NHS^{23,25,71-73}. Incluso se apunta que la exclusión de pacientes puede ser la causa de un inexplicable descenso en la incidencia de enfermedad coronaria cuando se utiliza como fuente de información las bases de datos de historias clínicas⁷³.

La perspectiva de los pacientes

Algunos trabajos sugieren que los pacientes deberían incrementar su influencia en la evaluación de la calidad de la atención que reciben y en las retribuciones de sus médicos, aunque reconocen que hay poca información para valorar si este tipo de medidas contribuiría o no a la mejora de la calidad de la atención⁷⁴. En todo caso, los trabajos actuales muestran que no hay relación entre los juicios de los pacientes sobre la calidad de los centros sanitarios y las puntuaciones de estos centros en los indicadores de calidad⁷⁵. Mayor acuerdo parece haber (en la literatura médica anglosajona, no en la realidad española) en que los pacientes tienen derecho a conocer los sistemas de pago a sus médicos, especialmente si los esquemas utilizados contienen incentivos para realizar o no realizar determinadas actuaciones⁷⁶, y ello pese a que los médicos eluden tratar este tema ante preguntas directas de los pacientes, y éstos pueden experimentar cierta incomodidad o desconfianza al conocer algunos de los incentivos por los que cobran sus médicos⁷⁷.

Impacto sobre el profesionalismo y la motivación interna de los médicos

Uno de los efectos paradójicos de los incentivos financieros es el posible impacto negativo sobre la moral y la motivación de los profesionales. Muchas actuaciones profesionales están intrínsecamente motivadas y hay evidencias de que este tipo de motivación puede ser socavado por las retribuciones externas⁷⁸, especialmente si crean conflictos internos. Se trata de un aspecto importante porque la motivación intrínseca se considera tradicionalmente un atributo esencial de la excelencia profesional⁷⁹.

La calidad es un problema del «sistema»

Los marcos conceptuales actuales tienden a definir los problemas de calidad como fallos del «sistema», esto es, del conjunto de la organización antes que de cada uno de los individuos que la integran. El hecho de enfocar los incentivos hacia los médicos individuales parece obviar este componente de sistema y, sobre todo, se asume que la calidad de la atención sanitaria deriva exclusivamente de las decisiones clínicas individuales y se apuesta por que las actuaciones financieras sobre los médicos pueden mejorar estas decisiones⁸⁰. Obviamente, no se trata de negar el peso de las decisiones clínicas individuales en la calidad de la atención, pero hay evidencias de que la calidad asistencial esta muy influida por la organización^{81,82} y, por tanto, los directivos de estas organizaciones no deberían descuidar su responsabilidad como «organización» escudándose en indicadores adecuados y suficiente dinero para incentivos. Estos elementos, incluso si están bien diseñados, tienen pocas posibilidades de producir grandes mejoras si no van acompañados de estrategias organizativas globales, capacidad técnica (inteligencia organizacional) y liderazgo gestor y clínico para conducir la organización. La Veterans Health Administration, una organización de carácter público que constituye el sistema de atención integrada más importante de Estados Unidos (con más de 100 hospitales, presta atención a unos 5 millones de ex militares y sus familias), suele ponerse como ejemplo de la importancia de las estrategias organizacionales⁸³.

Hay aspectos obvios que requieren abordajes organizativos. El rediseño de los procesos de atención basándose en las mejores prácticas, la incorporación inteligente de tecnologías de la información y soporte a la toma de decisiones, la coordinación de la atención de los pacientes (p. ej., qué procesos deben concentrarse en algunos centros para optimizar la relación volumen-calidad) o el desarrollo de prioridades y políticas de atención sobre las que incorporar las medidas de calidad y eficiencia. La importancia de los aspectos de sistema es aún más obvia, incluso cuando se requieren cambios organizativos para mejorar la atención y la sensibilidad a las necesidades de los pacientes o, simplemente, para incrementar la transparencia del sistema, una condición que va unida a los estudios evaluativos y a las propuestas de mejora.

Recomendaciones

El «pago por calidad» se está introduciendo con fuerza en los sistemas sanitarios de los países desarrollados, tanto en los que se basan en el mercado y esquemas retributivos básicos de pago por acto, como en

los sistemas públicos bajo esquemas básicos de capitación o salario, y tanto para el reembolso a hospitales u organizaciones integradas como para retribuir médicos individuales. Pese a sus limitaciones, y las escasas evidencias sobre su efectividad y capacidad para mejorar la calidad de la asistencia, el P4P ha venido para quedarse, aunque no necesariamente en las modalidades utilizadas actualmente. De hecho, una de las fortalezas del P4P es su plasticidad para ser modificado (en tipo de incentivos, cantidades, indicadores, criterios, etc.) y, por tanto, para intentar un realineamiento continuo entre retribución y mejora de calidad.

Los resultados de la investigación en servicios sanitarios suelen ser ambiguos, y pasar desde esta literatura médica a las recomendaciones para un sistema sanitario concreto siempre tiene, se haga o no explícito, un alto componente interpretativo. Tentativamente, los elementos de mayor interés para el SNS podrían ser:

– Los sistemas de indicadores utilizados mayoritariamente en el SNS, que fundamentalmente cargan contra los medicamentos de alto precio en atención primaria y contra la duración de la estancia y el rendimiento de quirófano en atención especializada, no parecen conceptualmente asociados a mejores resultados para los pacientes. Se requiere un importante esfuerzo para desarrollar nuevos indicadores que realmente conduzcan a una mejora de la calidad de la atención.

– Estos nuevos indicadores deben abordar, en alguna medida, la información sobre desigualdades. Construir separadamente indicadores para ambos sexos, para diferentes niveles socioeconómicos y para algunos grupos específicos puede ser el primer paso en este sentido.

– El desarrollo de indicadores más finos que los actuales estimulará las actuaciones de los agentes con intereses en el sector. Cabe esperar presiones de las sociedades de especialistas, de la industria farmacéutica y de las organizaciones de pacientes por incorporar indicadores que sitúen «su» medicamento, especialidad o enfermedad en el centro de los esquemas de P4P. Se apunta que la fijación de los puntos de corte de los indicadores en los valores recomendados por la GPC podría disparar de forma importante el gasto farmacéutico. Mantener la regla de fijar indicadores sólo para cuidados de alto valor clínico y, prioritariamente, en poblaciones de alto riesgo, será un elemento crítico para la efectividad y la eficiencia del pago por calidad.

– Seguir las reglas básicas de la teoría de incentivos es una recomendación obvia. Menos obvia, pero probablemente más importante, será establecer y dar importancia a los «incentivos» no económicos (a todo lo que favorece la motivación intrínseca y la satisfacción por el trabajo bien hecho) y el refuerzo del profesionalismo.

– Es importante situar los incentivos de forma que se favorezca la coordinación e integración de la atención. Pagar a los médicos de atención primaria por no hacer (no derivar, no prescribir, no pedir pruebas) y a los especialistas por hacer cuanto más mejor (más visitas, más pruebas, más ingresos, más intervenciones) puede ser contradictorio.

– El paternalismo inherente a los sistemas de indicadores (que establecen cómo deben tratarse todos los pacientes más allá de sus preferencias), necesita contrapesarse por la influencia de los pacientes. La debilidad actual de este contrapeso en el SNS sugiere la conveniencia de indicadores poco estrictos y/o incentivos poco potentes.

– El desarrollo de nuevos indicadores, su evaluación y, en general, el conjunto de las estrategias de P4P requieren mucho más análisis e información que la que está en condiciones de aportar la extraordinariamente débil investigación en servicios sanitarios española. No sólo se requieren más estructuras «inteligentes» (que, a su vez, costaría mucho formar), sino muchas más facilidades para el acceso a la información, esto es, mayor transparencia.

En todo caso, es importante recordar que los indicadores y los incentivos asociados no son el único instrumento de mejora de calidad disponible en el sistema sanitario. Es más, probablemente se trata de un instrumento que, como muestran las experiencias de la Veterans Administration o del propio National Health Service británico, sólo funciona adecuadamente formando parte de estrategias organizativas globales.

Agradecimientos y eximentes

Enrique Bernal, Beatriz González, Ricard Meneu, Vicent Ortún y Jaume Puig revisaron las versiones preliminares de este manuscrito y nos dieron sabios consejos que casi nunca seguimos, aunque siempre nos fueron útiles. Procede –y mucho– la eximente habitual. Las instituciones en que trabajan los autores no comparten necesariamente las opiniones contenidas en este trabajo.

Bibliografía

1. Robinson JC. Theory and practice in the design of physician payment incentives. *Milbank Q.* 2001;79:149-77.
2. Conrad DA, Christianson JB. Penetrating the «black box»: financial incentives for enhancing the quality of physician services. *Med Care Res Rev.* 2004;61 3 Suppl:37-68.
3. Rosenthal MB, Landon BE, Normand SL, Frank RG, Ahmad TS, Epstein AM. Employers' use of value-based purchasing strategies. *JAMA.* 2007;298:2281-8.
4. Epstein AM. Pay for performance at the tipping point. *N Engl J Med.* 2007;356:515-7.

5. Roland M. Linking physician pay to quality of care: a major experiment in the United Kingdom. *N Engl J Med.* 2004;351:1448-54.
6. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med.* 2006;355:375-84.
7. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med.* 2007;357:181-90.
8. Pink GH, Brown AD, Studer ML, Reiter KL, Leatt P. Pay-for-performance in publicly financed healthcare: some international experience and considerations for Canada. *Healthc Pap.* 2006;6:8-26.
9. Scott IA. Pay for performance in health care: strategic issues for Australian experiments. *Med J Aust.* 2007;187:31-5.
10. Bernal E, Meneu R, Pradas F, Peiró S, Ridao M. Sistemas de incentivos en atención primaria. *Rev Val Med Fam.* 1999;6:1-12.
11. Gené Badia J, Gallo de Puelles P. Retribución variable vinculada a la calidad asistencial. *Aten Primaria.* 2004;34:198-201.
12. Martín JJ. Motivación, incentivos y retribuciones de los médicos de atención primaria en el Sistema Nacional de Salud. *Rev Adm Sanit.* 2005;3:111-30.
13. Díaz Ojeda E, Navarro P, Prados S. Sistema Nacional de Salud: la retribución variable. *Rev Adm Sanit.* 2005;3:205-35.
14. Gérvas J, Ortún V, Palomo L, Ripoll MA. Seminario de Innovación en Atención Primaria 2007. Incentivos en atención primaria: de la contención del gasto a la salud de la población. *Rev Esp Salud Pública.* 2007;81:589-96.
15. Ortún Rubio V. Los incentivos para la mejora de la calidad en los servicios de salud. *Rev Calidad Asistencial.* 2007;22:1-6.
16. Gené-Badia J, Escaramis-Babiano G, Sans-Corrales M, Sampietro-Colom L, Aguado-Menguy F, Cabezas-Peña C, et al. Impact of economic incentives on quality of professional life and on end-user satisfaction in primary care. *Health Policy.* 2007;80:2-10.
17. Plaza Tesías A, Zara Yahni C, Guarga Rojas A, Farrés Quesada J. Resultado de la aplicación del benchmarking en los equipos de atención primaria de Barcelona. *Aten Primaria.* 2005;35:122-7.
18. Giuffrida A, Gosden T, Forland F, Kristiansen IS, Sergison M, Leese B, Pedersen L, Sutton M. Target payments in primary care: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 2000;3:CD000531.
19. Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS. Strategies to support quality-based purchasing: a review of the evidence. Rockville: Agency for Healthcare Research and Quality; 2004. Technical review 10; AHRQ publication 04-0057.
20. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145:265-72.
21. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev.* 2006;63:135-57.
22. Dudley RA. Pay-for-performance research: how to learn what clinicians and policy makers need to know. *JAMA.* 2005;294:1821-3.
23. Simpson CR, Hannaford PC, Lefevre K, Williams D. Effect of the UK incentive-based contract on the management of patients with stroke in primary care. *Stroke.* 2006;37:2354-60.
24. Kmietowicz Z. New GP contract: modernisation or miscalculation? *BMJ.* 2006;333:1192.
25. Sanderson AA. Pay-for-performance programs in the United Kingdom. *N Engl J Med.* 2006;355:1832.
26. Gené Badia J. Basta de «clickar» casillas. *Aten Primaria.* 2007;39:169-70.
27. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood).* 2004;23:127-41.
28. Fisher ES. Paying for performance: risks and recommendations. *N Engl J Med.* 2006;355:1845-7.
29. Bokhour BG, Burgess JF, Hook JM, White B, Berlowitz D, Guldin MR, et al. Incentive implementation in physician practices: a qualitative study of practice executive perspectives on pay for performance. *Med Care Res Rev.* 2006;63 Suppl 1:73-95.
30. Rowe JW. Pay-for-performance and accountability: related themes in improving health care. *Ann Intern Med.* 2006;145:695-9.
31. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA.* 2007;297:740-4.
32. Peiró Moreno S. De la gestión de lo complementario a la gestión integral de la atención de salud: gestión de enfermedades e indicadores de actividad. En: Ortún V, editor. *Gestión clínica y sanitaria: de la práctica academia, ida y vuelta.* Barcelona: Masson; 2003. p. 17-87.
33. Cooper BW. Public reporting and pay for performance. *N Engl J Med.* 2007;356:1783-4.
34. Wennberg JE, O'Connor AM, Collins ED, Weinstein JN. Extending the P4P agenda (I): how Medicare can improve patient decision making and reduce unnecessary care. *Health Aff (Millwood).* 2007;26:1564-74.
35. Wennberg JE, Fisher ES, Skinner JS, Bronner KK. Extending the P4P agenda (II): how Medicare can reduce waste and improve the care of the chronically ill. *Health Aff (Millwood).* 2007;26:1575-85.
36. Gosfield AG, Reinertsen JL. Paying physicians for high-quality care. *N Engl J Med.* 2004;350:1910.
37. Bierman AS, Clark JP. Performance measurement and equity. *BMJ.* 2007;334:1333-4.
38. Millett C, Saxena S, Ng A, Mainous A 3rd, Majeed A. Socio-economic status, ethnicity and diabetes management: an analysis of time trends using the health survey for England. *J Public Health (Oxf).* 2007;29:413-9.
39. Hippisley-Cox J, O'Hanlon S, Coupland C. Association of deprivation, ethnicity, and sex with quality indicators for diabetes: population based survey of 53,000 patients in primary care. *BMJ.* 2004;329:1267-9.
40. McLean G, Sutton M, Guthrie B. Deprivation and quality of primary care services: evidence for persistence of the inverse care law from the UK Quality and Outcomes Framework. *J Epidemiol Community Health.* 2006;60:917-22.
41. McLean G, Guthrie B, Sutton M. Differences in the quality of primary medical care for CVD and diabetes across the NHS: evidence from the quality and outcomes framework. *BMC Health Serv Res.* 2007;7:74.
42. Guthrie B, McLean G, Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. *Br J Gen Pract.* 2006;56:836-41.
43. Heath I, Hippisley-Cox J, Smeeth L. Measuring performance and missing the point? *BMJ.* 2007;335:1075-6.
44. Hayward RA. Performance measurement in search of a path. *N Engl J Med.* 2007;356:951-3.
45. Yarnall KS, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? *Am J Public Health.* 2003;93:635-41.

46. Østbye T, Yarnall KS, Krause KM, Pollak KI, Gradison M, Michener JL. Is there time for management of patients with chronic diseases in primary care? *Ann Fam Med*. 2005;3:209-14.
47. Hippisley-Cox J, Pringle M. Comorbidity of Diseases in the New General Medical Services Contract for General Practitioners: Analysis of QRESEARCH Data. Nottingham: QRESEARCH; 2005. Disponible en: www.qresearch.org/Public_Documents/DataValidation/Comorbidity%20of%20diseases%20in%20the%20new%20GMS%20contract%20for%20GPs.pdf
48. Ganz DA, Wenger NS, Roth CP, Kamberg CJ, Chang JT, MacLean CH, et al. The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences? *Med Care*. 2007;45:8-18.
49. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293:1239-44.
50. Brown C, Lilford R. Cross sectional study of performance indicators for English Primary Care Trusts: testing construct validity and identifying explanatory variables. *BMC Health Serv Res*. 2006;6:81.
51. Ganz DA, Wenger NS, Roth CP, Kamberg CJ, Chang JT, MacLean CH, et al. The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences? *Med Care*. 2007;45:8-18.
52. Glickman SW, Ou FS, DeLong ER, Roe MT, Lytle BL, Mulgund J, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007;297: 2373-80.
53. Fonarow GC, Abraham WT, Albert NM, Stough WG, Gheorghide M, Greenberg BH, et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. *JAMA*. 2007;297:61-70.
54. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006;296:2694-702. Erratum in: *JAMA*. 2007;297:700.
55. Horn SD. Performance measures and clinical outcomes. *JAMA*. 2006;296:2731-2.
56. Wang OJ, Wang Y, Lichtman JH, Bradley EH, Normand SL, Krumholz HM. «America's Best Hospitals» in the treatment of acute myocardial infarction. *Arch Intern Med*. 2007;167:1345-51.
57. Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA*. 2006;296:72-8.
58. Bottle A, Gnani S, Saxena S, Aylin P, Mainous AG, Majeed A. Association between quality of primary care and hospitalization for coronary heart disease in England: national cross-sectional study. *J Gen Intern Med*. 2007 [Epub ahead of print].
59. Downing A, Rudge G, Cheng Y, Tu YK, Keen J, Gilthorpe MS. Do the UK government's new Quality and Outcomes Framework (QOF) scores adequately measure primary care performance? A cross-sectional survey of routine healthcare data. *BMC Health Serv Res*. 2007;7:166.
60. Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff (Millwood)*. 2007;26:1104-10.
61. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. *JAMA*. 2004;291:2466-70.
62. Garber AM. Evidence-based guidelines as a foundation for performance incentives. *Health Aff (Millwood)*. 2005; 24:174-9.7
63. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298:1209-12.
64. Hayward RA, Hofer TP, Kerr EA, Krein SL. Quality improvement initiatives: issues in moving from diabetes guidelines to policy. *Diabetes Care*. 2004;27 Suppl 2:54-60.
65. Safford MM, Shewchuk R, Qu H, Williams JH, Estrada CA, Ovalle F, et al. Reasons for not intensifying medications: differentiating «clinical inertia» from appropriate care. *J Gen Intern Med*. 2007;22:1648-55.
66. Hayward RA, Hofer TP, Vijan S. Narrative review: lack of evidence for recommended low-density lipoprotein treatment targets: a solvable problem. *Ann Intern Med*. 2006;145:520-30.
67. McMahon LF Jr, Hofer TP, Hayward RA. Physician-level P4P-DOA? Can quality-based payment be resuscitated? *Am J Manag Care*. 2007;13:233-6.
68. Hayward RA. All-or-nothing treatment targets make bad performance measures. *Am J Manag Care*. 2007;13:126-8.
69. Hayward RA, Kent DM. Using clinical trial summary results to establish quality measures. *JAMA*. 2007;298:2741.
70. Vijan S, Hofer T, Hayward RA. Cost-utility analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *JAMA*. 2000; 283:889-96.
71. Vijan S, Hofer TP, Hayward RA. How often should patients with diabetes Be screened for retinopathy? *JAMA*. 2000;284:437-9.
72. Dicker A. Target tyranny. *J R Soc Med*. 2004;97:496-7.
73. Carey IM, Dewilde S, Harris T, Whincup PH, Cook DG. Spurious trends in coronary heart disease incidence: unintended consequences of the new GP contract? *Br J Gen Pract*. 2007;57:486-9.
74. Elwyn G, Buetow S, Hibbard J, Wensing M. Measuring quality through performance. Respecting the subjective: quality measurement from the patient's perspective. *BMJ*. 2007;335:1021-2.
75. Chang JT, Hays RD, Shekelle PG, MacLean CH, Solomon DH, Reuben DB, et al. Patients' global ratings of their health care are not associated with the technical quality of their care. *Ann Intern Med*. 2006;144:665-72. Erratum in: *Ann Intern Med*. 2006;145:635-6.
76. Pearson SD, Kleinman K, Rusinak D, Levinson W. A trial of disclosing physicians' financial incentives to patients. *Arch Intern Med*. 2006;166:623-8.
77. Pereira AG, Pearson SD. Patient attitudes toward physician financial incentives. *Arch Intern Med*. 2001;161:1313-7.
78. McDonald R, Harrison S, Checkland K, Campbell SM, Roland M. Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study. *BMJ*. 2007;334:1357.
79. Marshall M, Harrison S. It's about more than money: financial incentives and internal motivation. *Qual Saf Health Care*. 2005;14:4-5.
80. Gray BH. Individual incentives to fix organizational problems? *Med Care Res Rev*. 2004;61 3 Suppl:76-9.
81. Fleming B, Silver A, Oceppek-Weilikson K, Keller D. The relationship between organizational systems and clinical quality in diabetes care. *Am J Manag Care*. 2004;10:934-44.
82. Yano EM, Soban LM, Parkerton PH, Etzioni DA. Primary care practice organization influences colorectal cancer screening performance. *Health Serv Res*. 2007;42:1130-49.
83. Kerr EA, Fleming B. Making performance indicators work: experiences of US Veterans Health Administration. *BMJ*. 2007; 335:971-3.