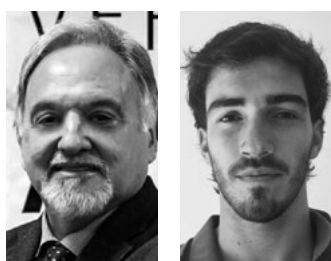


# Pago por desempeño y valor

## La clave es aprender de los errores y lograr una adecuada implementación



Por Carlos Alberto Díaz y Tobías Franzetti

El siguiente artículo plantea la necesidad de un análisis crítico sobre la implementación de los modelos de pago por desempeño (P4P), y cómo lograr un esquema costo-efectivo a partir de un aprendizaje de los errores previos para optimizar su efectividad. Sus limitaciones y la comparación con las experiencias internacionales

**E**l Pago por Desempeño (P4P), en inglés Pay for Performance, consiste en el uso de incentivos financieros orientados a promover mejoras en la efectividad, eficiencia y en la calidad de la atención médica brindada a los pacientes; además de mejorar resultados clínicos manteniendo sostenibilidad financiera del sistema de salud.

Este modelo forma parte de un conjunto de esquemas financieros denominados Modelos Alternativos de Pago (APM, por sus siglas en inglés Alternative Payment Models) diseñados para sustituir el sistema tradicional de Pago por Servicio Fee-for-Service (FFS) el cual recompensa el volumen de prestaciones

ofrecidas por encima de la calidad y la eficiencia del cuidado de la salud.

A partir de la década del 2000, la popularidad de los esquemas de P4P aumentó significativamente a nivel internacional, impulsados por la búsqueda de mecanismos de financiamiento basados en resultados y valor en salud, promovidos por organismos como los Centros de Servicios de Medicare y Medicaid (CMS) en los Estados Unidos.

El análisis del P4P resulta particularmente relevante en la actualidad, ya que representa uno de los intentos más sistemáticos por alinear los incentivos económicos con los objetivos clínicos y de salud pública. Compre-

\* **Díaz Carlos Alberto** es gerente médico del Sanatorio Sagrado Corazón. Profesor Titular de Universidad ISALUD.  
**Franzetti Tobías** es Médico. Residente de Gestión de Salud. Sanatorio Sagrado Corazón.

der su impacto permite evaluar si los incentivos financieros efectivamente mejoran los resultados clínicos, reducen costos y promueven la equidad, o si, por el contrario, generan conductas no deseadas como la selección de pacientes. (Mendelson, A., et al. 2017)<sup>1</sup>.

## Introducción a los tipos de P4P

Existen diversas modalidades dentro de los esquemas de P4P, las cuales difieren según los objetivos e indicadores utilizados para medir los resultados. Entre las principales se destacan las siguientes:

**Pago por calidad (Pay for Quality):** se basa en el cumplimiento de indicadores de calidad, que pueden incluir medidas de proceso, resultados clínicos, coordinación del cuidado y salud poblacional. Este enfoque busca recompensar la adhesión a prácticas basadas en la evidencia y la mejora continua en los estándares asistenciales.

**Pago por reporte (Pay for Reporting):** centra su objetivo en la transparencia y comunicación de datos de calidad. Representa un paso preliminar hacia los esquemas de pago por calidad, ya que incentiva a las instituciones a desarrollar capacidades de medición y registro antes de vincular los resultados a recompensas financieras.

**Pago por eficiencia (Pay for Efficiency):** orientado a incentivar la reducción de costos innecesarios, mediante indicadores como el gasto anual por paciente con enfermedades crónicas o los días de internación por paciente por año. Su propósito es promover un uso más racional de los recursos sanitarios sin comprometer la calidad del cuidado.

**Pago por valor (Pay for Value):** combina los enfoques anteriores, vinculando simultáneamente la calidad y la eficiencia. Recompensa la mejora de la calidad asistencial manteniendo los costos constantes o reduciéndolos, alineando así los incentivos con el paradigma de valor en salud (value-based care).

## P4P como política de salud en Estados Unidos a través del CMS

A pesar de la creciente popularidad de los esquemas de pago por desempeño (P4P), persisten interrogantes acerca de su efectividad real y preocupaciones

respecto de efectos adversos no intencionados, como la selección de pacientes o la priorización de indicadores fácilmente alcanzables. Analizar la experiencia acumulada en la implementación de programas de P4P en Estados Unidos resulta esencial para evaluar su impacto y diseñar estrategias que optimicen su aplicación en futuras políticas de salud.

Previo a la década del 2000, el modelo predominante era el Fee-for-Service (FFS), basado en el pago por volumen de servicios prestados, sin considerar la calidad ni los resultados clínicos obtenidos. Este esquema incentivaba la mayor producción de actos médicos, lo que contribuyó a mejorar la accesibilidad al sistema de salud en su momento, especialmente entre la población previamente no asegurada. Por lo que durante las décadas de 1950 y 1960, el FFS fue percibido incluso como un mecanismo de pago por calidad, asumiendo que los médicos por su formación y acreditación profesional garantizaban un estándar adecuado de atención. Sin embargo, a medida que el acceso se amplió y el gasto en salud aumentó, se evidenció que el modelo recompensaba el volumen, pero no necesariamente la efectividad, eficiencia ni valor de los servicios brindados.

El primer paso hacia la introducción de incentivos por calidad se produjo en 2003, cuando el Centers for Medicare & Medicaid Services (CMS) implementó el programa Hospital Quality Initiative (HQI). En este marco, los hospitales podían recibir un incremento del 0,4% en sus pagos anuales por el solo hecho de reportar públicamente sus indicadores de desempeño, independientemente de los resultados obtenidos. Este programa, denominado Reporting Hospital Quality Data for Annual Payment Update, buscaba fomentar la transparencia y disponibilidad de información sobre calidad de atención en áreas como infarto agudo de miocardio, neumonía, insuficiencia cardíaca, tiempos de administración de antibióticos, mortalidad y tasas de reingreso.

Posteriormente, el CMS avanzó hacia un modelo basado en resultados clínicos con el programa Premier Hospital Quality Incentive Demonstration (PHQID), considerado el primer ensayo formal de P4P en hospitales estadounidenses. Implementado en 2003,

involucró a 260 hospitales que recibían bonos financieros por alto desempeño y penalidades para los de menor rendimiento. Las métricas evaluadas incluían tanto procesos de atención (por ejemplo, administración de aspirina al ingreso o al alta, y uso oportuno de antibióticos) como resultados clínicos, tales como mortalidad y complicaciones en condiciones como infarto agudo de miocardio (AMI), neumonía adquirida en la comunidad (CAP), insuficiencia cardíaca congestiva (CHF) y cirugías coronarias u ortopédicas.

Finalmente, en 2012, el CMS evolucionó hacia un enfoque de pago por valor (Value-Based Purchasing) con la creación del programa Hospital Value-Based Purchasing (HVBP), sustentado en el Affordable Care Act de 2010. Este programa retiene un 2 % del pago base de cada hospital bajo Medicare y redistribuye esos fondos según el desempeño relativo. Los hospitales reciben bonificaciones si mejoran sus resultados o si superan su desempeño del año previo. El HVBP evalúa cuatro dimensiones clave:

- Resultados clínicos (mortalidad, complicaciones, reingresos, infarto, etc.).
- Experiencia del paciente (medida mediante la encuesta Hospital Consumer Assessment of Health-care Providers and Systems).
- Seguridad del paciente (infecciones asociadas al cuidado, caídas, eventos adversos).
- Eficiencia y reducción de costos (gasto por episodio de atención y uso adecuado de recursos).

## Experiencias internacionales de programas P4P

El Servicio Nacional de Salud (NHS) del Reino Unido implementó en 2008 el programa Advancing Quality, inspirado en el Hospital Quality Incentive Demonstration de Estados Unidos. El programa abarcó todos los hospitales del noreste de Inglaterra, cubriendo una población de aproximadamente 6,8 millones de personas.

Matt Sutton et al. (New England Journal of Medicine, 2012)<sup>2</sup> analizaron la mortalidad a 30 días en 134.435 pacientes hospitalizados por neumonía, insuficiencia cardíaca o infarto agudo de miocardio en

24 hospitales participantes del programa. Compararon la mortalidad durante los 18 meses previos y los 18 meses posteriores a la implementación del P4P. Se observó una reducción significativa de la mortalidad por neumonía (-1,9%), mientras que no hubo cambios estadísticamente significativos en los casos de insuficiencia cardíaca e infarto agudo de miocardio. Cabe destacar que, en comparación con los programas estadounidenses, los esquemas británicos presentan mayores incentivos económicos y una inversión institucional más robusta.

## Funcionamiento de los modelos P4P

Una revisión sistemática de 41 estudios sobre el desarrollo e implementación de indicadores en programas P4P mostró que los objetivos centrados en calidad clínica y experiencia del paciente se asociaron con mayor coordinación y comunicación entre proveedores, mientras que aquellos enfocados en productividad y eficiencia se vincularon con menor colaboración interdisciplinaria. Se recomienda que los indicadores combinen medidas de procesos y resultados clínicos, sean claros, factibles y basados en evidencia, y estén vinculados a resultados clínicamente relevantes. Asimismo, los incentivos financieros deben ser suficientemente grandes para motivar, pero no excesivos, a fin de evitar el gaming<sup>1</sup>. Las penalizaciones podrían ser más efectivas que las recompensas. Se observó que los niveles de calidad alcanzados suelen mantenerse tras la retirada de los incentivos, aunque los objetivos deben renovarse periódicamente para evitar estancamientos. (Kondo, Karli K et al. 2016)<sup>3</sup>

Un ensayo aleatorizado por conglomerados evaluó 29 organizaciones comunitarias dedicadas al tratamiento del consumo problemático en adolescentes, comparando 15 instituciones bajo esquema P4P con 14 controles. Ambas recibieron financiamiento similar para implementar el programa Adolescent Community Reinforcement Approach (A-CRA). El P4P resultó ser una estrategia costo-efectiva, con un aumento del 5 % en costos totales asociado a un

1. Aplica principios de diseño de juegos como puntos, recompensas y desafíos a actividades relacionadas con la salud para motivar a los médicos o los pacientes y mejorar los resultados

incremento del 325 % en pacientes tratados y días de abstinencia. Otros estudios reportaron aumentos de 16 % en costos, que se tradujeron en mayores tasas de vacunación antigripal (+9 %) y derivaciones a líneas de cesación tabáquica (+236 %), evidenciando que las mejoras superaron proporcionalmente el incremento de costos. Los incentivos dirigidos directamente a los proveedores, junto con pagos mensuales, se asociaron con mejores resultados que los pagos realizados a las organizaciones o de manera anual. (Garner, B.R., Lwin, A.K., Strickler, G.K. et al. 2018)<sup>4</sup>

El efecto de la implementación de un programa P4P en salud depende en gran medida del diseño del programa, incluyendo la estructura del incentivo financiero, la fuente de financiamiento, y los indicadores y objetivos que determinan la recompensa. Los programas pueden enfocarse en reducir costos (disminuyendo tiempos de estancia, readmisiones o uso de tecnologías terapéuticas), en la reducción de mortalidad o en el diseño de procesos que incrementen la calidad de la atención. Sin embargo, debe considerarse que la implementación de P4P conlleva gastos administrativos adicionales, como recolección de datos, auditorías y desarrollo/mantenimiento de medidas de mejora.

El Premier Hospital Quality Incentive Demonstration (PHQID) del CMS otorgó incentivos financieros por mejoras en el desempeño hospitalario, con la premisa de que los desperdicios podrían reducirse mediante la optimización de recursos y eficiencia en la atención. Sin embargo, no existen estudios publicados que evalúen el impacto de P4P sobre el uso de recursos o la eficiencia de costos de manera directa, dado que los programas típicos de P4P no incluyen medidas explícitas de eficiencia. Pese a esto, los modelos P4P demostraron mejoras en la provisión confiable de cuidados adecuados. (RAND, Analysis of Hospital Pay for Performance)<sup>5</sup>

## **Puntos clave para el diseño de un programa de Pago por Desempeño (P4P): aprendizajes de éxitos y errores**

El diseño de los programas de P4P depende de múltiples factores: el indicador elegido para medir

el desempeño, el monto asignado como incentivo, la modalidad de los incentivos (recompensas, penalizaciones o ambos), la frecuencia de los pagos y el tipo de establecimiento de salud y su cultura organizacional, así como la población de pacientes atendida.

## **Cómo se mide el desempeño**

Primero, se deben definir las dimensiones de desempeño que el programa recompensará. En salud, estas pueden incluir:

- Resultados clínicos (QALYs): Un ejemplo clásico es la medición de cambios en mortalidad, pero es difícil de medir e implica cambios a largo plazo. Por ejemplo, Ubel et al., 2003<sup>6</sup> y Nahra et al., 2006<sup>7</sup> analizaron el programa de Blue Cross Blue Shield en el Hospital de Michigan, observando un aumento en el número de pacientes que recibieron tratamiento adecuado para infarto agudo de miocardio e insuficiencia cardíaca congestiva. El estudio de Nahra et al. reportó una ganancia de 733 a 1.701 QALYs, mientras que Ubel et al. estimó un costo de \$12.967 a \$30.081 por QALY, un rango considerado costoefectivo.
- Calidad de procesos clínicos: se ha convertido en la manera más práctica de evaluar los resultados clínicos, por ejemplo, la cantidad de mamografías o colonoscopias realizadas. (El riesgo es basar indicadores en opiniones de expertos en lugar de evidencia de ensayos aleatorizados).
- Seguridad del paciente: Aumento en el cumplimiento de lavado de manos.
- Accesibilidad: tiempos de espera, cobertura de servicios.
- Satisfacción de pacientes: generalmente medida mediante encuestas.
- Eficiencia de costos: costo de generar un QALY adicional.
- Efectividad de costos: evaluación de intervenciones alternativas que producen resultados similares, como la reducción de estudios diagnósticos innecesarios.
- Adherencia a prácticas basadas en evidencia.
- Productividad: Eficiencia con la que los recursos se transforman en resultados. Se recompensa no

solo la cantidad, sino el resultado y la calidad de atención

- Adopción de tecnología: implementación de historias clínicas electrónicas
- Reportes por desempeño: La publicación de indicadores, independientemente de los resultados obtenidos, busca promover la transparencia y constituye un primer paso esencial para la recolección de datos. La importancia de expandir este modelo radica en su capacidad para servir como base comparativa frente a esquemas de P4P. Un estudio comparó 252 hospitales participantes en el PHQID con 3.363 hospitales control que únicamente formaban parte del programa de Pay for Reporting. En una cohorte de más de 6 millones de pacientes con infarto agudo de miocardio (AMI), insuficiencia cardíaca congestiva (CHF) o neumonía atendidos entre 2003 y 2009, no se encontró evidencia de una reducción significativa en la mortalidad a 30 días atribuible al programa de P4P. Las tasas de mortalidad iniciales fueron similares (12,33 % vs. 12,40 %) y, tras seis años, se mantuvieron prácticamente iguales (11,82 % vs. 11,74 %) (Jha AK, et al., 2012, NEJM)<sup>8</sup>. Por su parte, Lindenauer et al. (2007, NEJM)<sup>9</sup> evaluaron 613 hospitales y analizaron la adherencia a 10 indicadores individuales y 4 compuestos de calidad clínica. Al comparar hospitales del PHQID con aquellos que participaban únicamente en el programa de reporte, observaron mejoras en el rango de 2,6 % a 4,1 % en los indicadores de desempeño durante un período de dos años. Los autores concluyeron que los hospitales involucrados en ambos programas presentaron mayores mejoras relativas, aunque destacaron que no se evaluó la costoefectividad del programa.
- Participación en actividades de mejora de procesos

El CMS distingue dos componentes principales para el cálculo de incentivos:

- Logros de calidad: Evalúa el nivel absoluto de desempeño, considerando los resultados de los indicadores del hospital en comparación con benchmarks o metas nacionales.

Mejoras de calidad: Evalúa el progreso respecto a años anteriores, midiendo cuánto ha mejorado el hospital en relación con el desempeño de un período previo.

Se ha demostrado que los logros de calidad suelen ser más efectivos que los pagos por mejoras de calidad. (Mathes, Tim et al. 2019)<sup>10</sup>

## Selección y asignación de indicadores

Al elegir indicadores específicos, es crucial considerar:

- Validez: deben reflejar fielmente la dimensión que se pretende medir.
- Fiabilidad: deben mostrar de manera consistente los cambios en desempeño clínico.
- Relevancia: deben medir aspectos significativos de cada dimensión.
- Especificidad: cada indicador debe ser específico a la dimensión evaluada.
- Controlabilidad: deben estar bajo el control de la organización que implementa el P4P.
- Accionabilidad: deben medir factores sobre los cuales la organización pueda intervenir para lograr algo positivo.
- Eficiencia: el conjunto de indicadores debe ser reducido, evitando exceso de datos que aumenten costos administrativos sin agregar valor.
- Efectividad de costos: el beneficio de los indicadores debe superar los costos de recolección y monitoreo.

## Financiamiento de los programas P4P

Según el Institute of Medicine, Board on Health Care Services (2007)<sup>11</sup>, existen tres formas de financiar un P4P:

- Pagos existentes (Existing payments): redistribución de pagos ya establecidos, beneficiando al financiador, pero penalizando a proveedores de bajo desempeño.
- Ahorros generados (Generated savings): utilizar mejoras en eficiencia y calidad que reduzcan costos para financiar los incentivos.
- Dinero nuevo (New money): asignación de fondos

adicionales al presupuesto existente, necesaria para cubrir gastos administrativos, capacitación, tecnología e inversión inicial. Sin embargo, esto plantea interrogantes sobre la eficiencia costobeneficio de las medidas implementadas y su justificación respecto al aumento del presupuesto.

## Implementación de incentivos financieros

Formas de distribuir los incentivos:

- Bonos y retenciones: Pagados al final del período de medición, generalmente de manera anual. Por ejemplo, el programa Blue Cross of California Preferred Provider Organization Rewarding Results otorgó un bono de \$5.000 a los clínicos en función de su desempeño (Young et al., 2007)<sup>12</sup>. El programa Excellus/Rochester (New York) Individual Practice Association Rewarding Results generaba una retención del 10 % y, en función del desempeño, devolvía entre el 50 % y el 150 %.
- Penalidades: Reducción de pagos si no se cumple el indicador de desempeño establecido. El programa Medicare Premier Hospital Quality Incentive Demonstration impuso una penalidad del 1 % a los hospitales que se desempeñaran por debajo del noveno decil de calidad de referencia y del 2 % si estaban por debajo del décimo decil (CMS, 2009b).
- Ajustes de esquemas de FeeforService: Ajuste de las tarifas fijas según el cumplimiento del indicador seleccionado. Por ejemplo, el programa Blue Cross Blue Shield of Michigan Rewarding Results permitía a los hospitales ganar hasta un 4 % por alcanzar los umbrales absolutos de desempeño en medidas de calidad seleccionadas (Young et al., 2007)<sup>12</sup>.
- Pago por miembro: En entornos de pagos per cápita, se puede otorgar un adicional por cada miembro extra adquirido. El programa Bridges to Excellence Rewarding Results pagaba un bono anual de \$100 por paciente para atención de diabetes y \$160 para atención cardíaca, según el reconocimiento de desempeño del National Committee for Quality Assurance (Young et al., 2007)<sup>12</sup>.
- Actualización de pago diferencial: Incentivo incremental para quienes se desempeñaron mejor durante el período analizado. En el programa Medicare Reporting Hospital Quality Data for Annual Payment Update, los hospitales que no reportaron las medidas de calidad designadas recibieron una reducción del 0,4 % (posteriormente incrementada al 2 %) en la actualización anual de sus pagos (CMS, 2009a).
- Pago por reporte o participación: Incentivos por participar en actividades de mejora continua, desarrollo de proyectos de mejora o capacitaciones. El programa Medicare Physician Quality Reporting Initiative pagaba a los médicos que reportaban exitosamente el 2 % de sus cargos permitidos por Medicare.
- Rechazo de pago por baja performance: Los financiadores pueden optar por no cubrir ciertos pagos asociados a indicadores de baja calidad. Desde 2008, Medicare dejó de pagar por costos adicionales relacionados con ocho eventos prevenibles, incluidos transfusiones con tipo de sangre incorrecto, úlceras por presión e infecciones intrahospitalarias específicas.

## Magnitud de los incentivos

Se debe analizar tanto la magnitud como la frecuencia del incentivo. Este debe estar relacionado con el costo que implica para el proveedor la intervención que el financiador busca recompensar. Por ejemplo, aumentar la cantidad de recetas no genera un costo incremental significativo, mientras que la implementación de historias clínicas electrónicas implica costos de instalación y mantenimiento.

Con excepción del Reino Unido, los programas generalmente comienzan con incentivos bajos, debido a preocupaciones sobre fiabilidad, validez, calidad, capacidad de recolección de datos y la posible controversia derivada de la disparidad de pagos. La mayoría de los esquemas de P4P en EE. UU. representa menos del 5 % de los ingresos totales de los médicos. Una revisión sugiere que incentivos de aproximadamente el 5 % de los ingresos totales del médico son suficientes para atraer “atención significativa” por parte de los profesionales (Young et al.,

2007)<sup>12</sup>. Además, limitar la magnitud evita conductas de gaming, que pueden anular el efecto deseado del P4P.

Otro factor relevante es la frecuencia de los incentivos. Aunque los incentivos anuales son comunes, aumentar la frecuencia proporciona mayor visibilidad al programa y puede tener un impacto más fuerte en la conducta de los participantes. No obstante, incrementar la frecuencia puede diluir los efectos individuales de cada pago y generar costos administrativos adicionales.

Finalmente, se debe decidir si el incentivo se distribuirá a nivel organizacional, por servicio o directamente al clínico, considerando la posible conducta de free rider, en la que algunos médicos se benefician del esfuerzo de sus colegas.

## Limitaciones de los modelos P4P

### Falta de indicadores de desempeño válidos, relevantes y confiables

Los resultados en salud están influenciados por múltiples factores: la participación de diversos clínicos y especialistas, la fragmentación del sistema sanitario, las políticas públicas, y factores ambientales y socioeconómicos. Por ello, cuantificar los resultados atribuibles a un solo individuo o institución resulta complejo. La relación entre los procesos de atención y los resultados finales es difícil de aislar, y muchas áreas médicas carecen aún de consenso respecto a las mejores prácticas clínicas. En consecuencia, la capacidad del financiador para reflejar el verdadero desempeño de los proveedores mediante los indicadores seleccionados es significativamente limitada.

El gasto en salud, comúnmente utilizado como indicador de eficiencia, suele ser una medida inadecuada, ya que está sujeto a una alta variabilidad dependiente de factores externos a la política pública implementada. Por ello, aislar indicadores y atribuir variaciones a la acción de los clínicos o incluso de los centros de salud es metodológicamente complejo. Es importante destacar que los financiadores tienden a basar los programas de P4P en los

indicadores disponibles, más que en los más relevantes u óptimos, debido a limitaciones de acceso a datos. Paradójicamente, mientras la práctica clínica se sustenta en la evidencia científica, la selección de indicadores de desempeño muchas veces no considera ni la significancia ni la magnitud de su asociación con los objetivos del programa, ni evalúa adecuadamente los factores externos que podrían influir sobre ellos.

Un ejemplo paradigmático es el uso de la mortalidad infantil como indicador global de eficiencia en salud entre países. Si se observa el gasto en salud como porcentaje del PBI, Argentina 9,86% y Bolivia 8,43% (Banco Mundial, 2022), podría inferirse erróneamente que la mayor mortalidad infantil de Bolivia (20 por 1000 nacidos vivos) frente a la de Argentina (8,2 por 1000 nacidos vivos) refleja una gestión deficiente o un uso ineficiente de los recursos. Sin embargo, factores externos, como la elevada altitud, afectan significativamente la salud materno-fetal al incrementar la mortalidad intrauterina y la incidencia de preeclampsia e hipertensión gestacional, reduciendo el peso al nacer (Keyes et al., 2003; Soria et al., 2013)<sup>13</sup>. Por lo tanto, juzgar la eficiencia del sistema boliviano únicamente mediante la mortalidad infantil resulta injustificado y metodológicamente sesgado.

Este ejemplo evidencia la necesidad de revisar y rotar periódicamente los conjuntos de indicadores empleados en los programas de P4P, evitando que las organizaciones proveedoras orienten sus esfuerzos exclusivamente hacia un solo indicador a expensas de los objetivos integrales del programa.

### Limitaciones económicas y de costo-efectividad

La falta de evidencia sobre la costo-efectividad constituye una barrera central para la adopción de programas P4P. Estos esquemas implican costos considerables: la medición y el reporte requieren inversión administrativa sostenida, además de los costos asociados a la implementación de mejoras en el desempeño, la capacitación del personal y el desarrollo de infraestructura tecnológica. Asimismo, los financiadores deben afrontar gastos derivados de la

gestión del programa, la recolección y evaluación de datos, y la administración de los incentivos.

En numerosos estudios, el incremento en el desempeño no justificó los costos adicionales de implementación, y la costo-efectividad no pudo ser demostrada de manera consistente. Esto es especialmente relevante cuando los beneficios esperados son de largo plazo. Por tanto, resulta esencial definir con precisión si los costos incrementales de los programas P4P serán absorbidos por el financiador o trasladados al proveedor.

El programa Medicare Coordinated Care Demonstration pagó a 15 organizaciones de manejo de enfermedades honorarios mensuales de gestión, promediando \$253 por beneficiario, con el objetivo de mejorar la coordinación de la atención de enfermedades crónicas, reducir costos y mantener o mejorar la calidad de la atención. Todas las organizaciones asignaron coordinadores de enfermería para los pacientes. Sin embargo, ninguno de los programas produjo reducciones significativas en costos frente a grupos control.

De manera complementaria, una revisión sistemática de Mathes et al. (2019, Cochrane Database of Systematic Reviews)<sup>14</sup> analizó 27 estudios que comparaban programas de P4P con otros esquemas de pago, abarcando entre 10 y 4.267 instituciones. Los autores concluyeron que la evidencia sobre el impacto del P4P en la calidad, los resultados clínicos y el uso de recursos es incierta, observándose mejoras modestas y de corto plazo que no lograron sostenerse a largo plazo.

## Consecuencias no deseadas

Entre las consecuencias no intencionadas más relevantes se encuentran el rechazo de pacientes de mayor complejidad por parte de los proveedores, debido al impacto negativo que estos casos pueden tener sobre los indicadores evaluados (Epstein et al., 2004)<sup>15</sup>. Paradójicamente, son estos pacientes quienes más se beneficiarían de mejoras en la calidad asistencial. Una posible solución consiste en ajustar los incentivos por riesgo, considerando la complejidad del caso como una variable moderadora del indicador.

Otra consecuencia es la profundización de las desigualdades en salud. Los proveedores que atien-

den poblaciones de altos ingresos suelen contar con mayor infraestructura y capacidad tecnológica que aquellos que atienden sectores vulnerables, lo que limita su posibilidad de alcanzar los objetivos exigidos (Felt-Lisk et al., 2007)<sup>16</sup>.

Además, pueden surgir efectos adversos cuando los programas no consideran diferencias poblacionales, asignan montos excesivos al riesgo financiero o fomentan conductas contrarias a la práctica clínica basada en evidencia. Entre ellas destacan:

- Sobreutilización de servicios.
- Manipulación de datos, como la exclusión de pacientes de alto riesgo (“gaming”).
- “Teaching to the test”, es decir, la concentración de esfuerzos solo en los indicadores recompensados, sin considerar los beneficios clínicos reales.
- Rechazo a pacientes complejos, para evitar penalizaciones en los indicadores.

Un caso ilustrativo es el programa PHQID, que, según Petersen et al. (2006)<sup>17</sup>, mostró una disminución en el desempeño de indicadores no incluidos en el esquema de incentivos. Asimismo, estudios de Drake, Cohen y Cohn (2007)<sup>18</sup> reportaron un sobreuso de antibióticos en patologías respiratorias, debido a un indicador que exigía su administración dentro de las cuatro horas del ingreso hospitalario, reportes concluyeron que los clínicos no podían establecer el diagnóstico de neumonía dentro de las cuatro horas de ventana estipuladas por el indicador, optando por prescribir antibióticos a una gran población de pacientes con síntomas respiratorios.

Un análisis comparativo entre el QOF del Reino Unido, los programas P4P estadounidenses y el P4P para diabetes (DM-P4P) de Taiwán mostró resultados heterogéneos. En el caso británico, se observaron mejoras sustanciales en la prescripción de anticonceptivos y medicación para diabetes, así como un aumento del diagnóstico de depresión, aunque sin cambios en la prescripción de antidepresivos. En Taiwán, en cambio, se detectó un fenómeno de “gaming”, ya que los médicos tendían a registrar en el programa a pacientes más sanos, generando una población beneficiaria artificialmente más saludable. En conjunto, no se hallaron mejoras sostenidas

en la calidad de la atención ni en los resultados clínicos. (Aaron Mendelson et al., 2017)<sup>19</sup>

## Conclusión

El desafío principal de los programas de pago por desempeño (P4P) consiste en alinearlos correctamente con otras intervenciones a nivel del médico individual, grupo médico, hospital, sistema de prestación, comunidad y políticas públicas. Es fundamental evitar incentivos que promuevan la selección de pacientes más acomodados, lo que podría inflar artificialmente los indicadores de calidad y aumentar las desigualdades en la atención. Asimismo, se debe garantizar que las instituciones que atienden a una mayor proporción de pacientes de bajos ingresos reciban financiamiento suficiente para competir eficazmente por los pagos por incentivos.

La revisión de todos los estudios analizados resalta la importancia y la creciente popularidad de los esquemas de P4P. Sin embargo, también evidencia que existe limitada evidencia científica que respalde su costo-efectividad, lo que plantea la necesidad de un análisis crítico sobre su implementación.

No se busca desestimar los programas de P4P, que con un diseño adecuado pueden representar el futuro de los modelos de pago en salud. La clave está en aprender de los errores y fracasos previos para optimizar su efectividad. A partir de esta premisa, el análisis se enfoca en cómo lograr un esquema de P4P costo-efectivo mediante una correcta implementación.

Para estructurar un programa de P4P, se recomienda partir de la teoría de análisis de sistemas multivariados, evaluando cómo diferentes variables influyen en los resultados y la incidencia conjunta sobre la variable de interés. En primer lugar, se deben definir claramente los objetivos del programa, priorizando aquellos que generen reducción de costos directa y a corto plazo para el financiador.

Una vez definidos los objetivos, se deben asignar indicadores de desempeño que reflejen la obtención de cada objetivo, asegurando que cumplan con todos los criterios de validación, fiabilidad, relevancia, especificidad, controlabilidad y acción. Este paso es crítico, pues muchos fracasos de los programas de P4P provienen de indicadores mal dise-


ñados o poco sensibles a los cambios conductuales inducidos por los incentivos.

Además de la conducta generada por los incentivos, es necesario identificar otros factores que puedan influir significativamente en los indicadores, agrupados en: determinantes sociales, estructura del sistema, conducta de los actores y políticas públicas. Esto permitirá evaluar la magnitud del efecto real de los incentivos sobre los indicadores, sin perder de vista las influencias externas.

Se debe diseñar un esquema de seguimiento que considere los costos administrativos y la facilidad de recolección de datos. Es recomendable mantener un número moderado de indicadores, ponderando su contribución al pago por desempeño.

La frecuencia ideal de pago podría ser semestral, preferiblemente intercalada con los períodos de aguinaldo, y el monto asociado debería reflejar la reducción de costos generada por los objetivos cumplidos, sin superar el 5% del sueldo del participante. Este esquema reduce los costos administrativos y mantiene la motivación activa de los participantes.

Ho, Leonard et al. (2025)<sup>20</sup> realizaron una revisión sistemática a uno y tres años del impacto del programa Quality and Outcomes Framework (QOF) en la calidad de la atención y los resultados en salud, tanto tras su introducción como después de su cancelación. Se incluyeron 30 estudios, mostrando un aumento significativo en 83 indicadores de calidad durante el primer año (+6,3% sobre la tendencia previa), aunque a los tres años el incremento fue solo del +0,7%. Tras la retirada de los incentivos, los indicadores específicos se redujeron en -10,7% al año y -12,8% a los tres años, evidenciando que las mejoras logradas tendían a revertirse. En promedio, las pérdidas de calidad tras la eliminación de los incentivos igualaron o superaron las ganancias iniciales.

Por último, se recomienda actualizar los objetivos del programa cada 2 a 4 años, o una vez alcanzadas las metas deseadas, manteniendo la supervisión constante para asegurar que los resultados obtenidos no disminuyan. Se debe considerar la posibilidad de incorporar objetivos previos al programa en la penalización de incentivos si los resultados decaen, o reincorporarlos al programa si es necesario, garantizando un aprendizaje continuo y sostenibilidad de los logros. 

## Bibliografía

1. Mendelson, Aaron et al. "The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review." *Annals of internal medicine* vol. 166,5 (2017): 341-353. doi:10.7326/M16-1881
2. Sutton, Matt, et al. "Reduced Mortality with Hospital Pay for Performance in England." *The New England Journal of Medicine*, vol. 367, no. 19, 2012, pp. 1821–1828. *New England Journal of Medicine*, <https://doi.org/10.1056/NEJMsa1114951>
3. Kondo, Karli K et al. "Implementation Processes and Pay for Performance in Healthcare: A Systematic Review." *Journal of general internal medicine* vol. 31 Suppl 1, Suppl 1 (2016): 61-9. doi:10.1007/s11606-015-3567-0
4. Garner, B.R., Lwin, A.K., Strickler, G.K. et al. Pay-for-performance as a cost-effective implementation strategy: results from a cluster randomized trial. *Implementation Sci* 13, 92 (2018). <https://doi.org/10.1186/s13012-018-0774-1>
5. [https://www.rand.org/pubs/technical\\_reports/TR562z12/analysis-of-hospital-pay-for-performance.html](https://www.rand.org/pubs/technical_reports/TR562z12/analysis-of-hospital-pay-for-performance.html)
6. Ubel, Peter A et al. "Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public." *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* vol. 12,6 (2003): 599-607. doi:10.1023/a:1025119931010
7. Nahra, Tammie & Reiter, Kristin & Hirth, Richard & Shermer, Janet & Wheeler, John. (2006). Cost-Effectiveness of Hospital Pay-for-Performance Incentives. *Medical care research and review : MCRR*. 63. 49S-72S. 10.1177/1077558705283629.
8. Jha, Ashish K., et al. "The Long-Term Effect of Premier Pay for Performance on Patient Outcomes." *The New England Journal of Medicine*, vol. 366, no. 17, 2012, pp. 1606–1615. *New England Journal of Medicine*, <https://doi.org/10.1056/NEJMsa1112351>
9. Lindenauer, Peter K., et al. "Public Reporting and Pay for Performance in Hospital Quality Improvement." *The New England Journal of Medicine*, vol. 356, no. 5, 2007, pp. 486–496. *New England Journal of Medicine*, <https://doi.org/10.1056/NEJMsa064964>
10. Mathes, Tim et al. "Pay for performance for hospitals." *The Cochrane database of systematic reviews* vol. 7,7 CD011156. 5 Jul. 2019, doi:10.1002/14651858.CD011156.pub2
11. Institute of Medicine (US). *Evidence-Based Medicine and the Changing Nature of Healthcare: 2007 IOM Annual Meeting Summary*. National Academies Press (US), 2008. doi:10.17226/12041
12. Young, Gary J., et al. «Lessons from the Rewarding Results Demonstrations.» *Health Care Financing Review*, vol. 29, no. 4, 2007, pp. 59–72. Centers for Medicare & Medicaid Services, <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HealthCareFinancingReview/downloads/07Fallpgs9.pdf>
13. Keyes, Linda E et al. "Intrauterine growth restriction, preeclampsia, and intrauterine mortality at high altitude in Bolivia." *Pediatric research* vol. 54,1 (2003): 20-5. doi:10.1203/01.PDR.0000069846.64389.DC
14. Mathes, Tim, et al. "Pay for Performance for Hospitals." *Cochrane Database of Systematic Reviews*, vol. 2019, no. 7, 2019, CD011156. <https://doi.org/10.1002/14651858.CD011156.pub2>
15. Epstein, Arnold M., et al. "Paying Physicians for High-Quality Care." *The New England Journal of Medicine*, vol. 350, no. 6, 2004, pp. 507–510. <https://doi.org/10.1056/NEJMp038205>
16. Felt-Lisk, Suzanne, et al. "Making Pay-for-Performance Work in Medicaid." *Health Affairs*, vol. 26, no. 4, 2007, pp. w516–w527. <https://doi.org/10.1377/hlthaff.26.4.w516>
17. Petersen, Laura A., et al. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine*, vol. 145, no. 4, 2006, pp. 265–272. <https://doi.org/10.7326/0003-4819-145-4-200608150-00006>
18. Drake, D.E., Cohen, A., & Cohn, J. (2007). National hospital antibiotic timing measures for pneumonia and antibiotic overuse. *Quality Management in Health Care*, 16(4), 322–327. <https://doi.org/10.1097/01.QMH.0000290744.34665.5b>
19. Mendelson, Aaron, et al. "The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review." *Annals of Internal Medicine*, vol. 166, no. 5, 7 Mar. 2017, pp. 341–353. <https://doi.org/10.7326/M16-1881>
20. Ho, Leonard et al. "Effect of UK Quality and Outcomes Framework pay-for-performance programme on quality of primary care: systematic review with quantitative synthesis." *BMJ (Clinical research ed.)* vol. 389 e083424. 25 Jun. 2025, doi:10.1136/bmj-2024-083424
21. *Pay for Performance in Health Care: Methods and Approaches* EDITED BY Jerry Cromwell, Michael G. Trisolini, Gregory C. Pope, Janet B. Mitchell, and Leslie M. Greenwald.