

DOCUMENTO RESUMEN OPERACIÓN FORMULA DE DESEMPEÑO

NUEVO MODELO DE REMUNERACIÓN:
PAGO POR DESEMPEÑO PRESUPUESTO
GLOBAL PROSPECTIVO POR
ACTIVIDAD FINAL

Alcaldía de Bogotá

DOCUMENTO RESUMEN OPERACIÓN FÓRMULA DE DESEMPEÑO

Nuevo modelo de remuneración: pago por desempeño

Presupuesto global prospectivo por actividad final

1. Definición

El nuevo modelo de remuneración (NMR) es un mecanismo de pago por desempeño (PPD) o resultados que busca mejorar la calidad de la atención en salud de la población de régimen subsidiado afiliada a Capital Salud, así como el equilibrio financiero del asegurador (Capital Salud) y el prestador (Subredes Integradas de Servicios de Salud ESE).

El NMR es, básicamente, un pago que incentiva la eficiencia en la utilización de recursos humanos y físicos del prestador y el cumplimiento en indicadores de calidad. Su principio es discriminar el valor pactado en un valor de referencia eficiente y un estímulo al desempeño: así no se excede el valor pactado, pero se lleva al prestador a buscar la eficiencia y garantizar la calidad de las atenciones. Dicho valor de referencia se calculó seleccionando el costo unitario más eficiente (menor costo-mayor producción) de todos los 22 hospitales que actualmente conforman las cuatro subredes ESE para cada actividad final. Este mecanismo de pago con incentivos se basa en los hallazgos realizados por la Organización para la Cooperación y el Desarrollo Económicos (OCDE), en relación con las características y los objetivos de los mecanismos de pago por desempeño encontrados a escala mundial, tanto para países en desarrollo como para países desarrollados. Se fundamenta en el concepto de presupuesto global prospectivo por actividad final (PGPAF). Un presupuesto global prospectivo consiste en que el asegurador y el prestador pactan de forma prospectiva un presupuesto (valor) global para pagar por una cantidad global de atenciones en salud (la

cantidad de atenciones entre un valor mínimo y uno máximo, o lo que se denomina *senda*), las cuales están medidas en actividades finales (consulta, cirugía y hospitalización). La esencia de este modelo es reducir los costos de transacción al simplificar los procesos de autorizaciones, facturación y auditoría.

Este mecanismo de pago con incentivos se basa en los hallazgos realizados por la OCDE (Cashin, Chi, Smith, Borowitz, & Thomson, 2014) y por Miller y Babiarz (2014), en relación con las características y los objetivos de los mecanismos de pago por desempeño encontrados a escala mundial, tanto para países en desarrollo (Miller et al., 2014), como para países desarrollados (Cashin et al., 2014).

2. Aspecto fundamentales en la operación del NMR PGPAFED

Las siguientes premisas representan las tres reglas de oro del NMR que permiten el equilibrio financiero entre el asegurador y el prestador:

2.1. Pacto de confianza VALOR GLOBAL = PRODUCCIÓN (Presupuesto Global)

Esto traduce que:

Si la subred cumple con las actividades dentro de una senda de producción pactada *ex ante* y una calidad mínima esperada, recibe el valor total del presupuesto global definido en el pacto inicial para dicha actividad. Para que esto sea una realidad matemática, la tarifa es dinámica o variable, y es determinada a partir de un valor de referencia y el grado de cumplimiento en el número de actividades pactadas y en la calidad esperada.

2.2. Disminución de costos de transacción (simplificación de procesos de autorizaciones, facturación y auditoría)

Las 4 subredes destinaron en la vigencia 2016 la suma de \$79.782.218.337 para realizar procesos administrativos relativos con la operación del contrato (costos de transacción entre el asegurador y el prestador, en procesos de autorizaciones, facturación y auditoría), suma que podría ser invertida en más y mejores atenciones en salud a la población. La disminución de costos de transacción es una meta fundamental del NMR.

Los sistemas de información de las cuatro subredes ya se encuentran conectados desde la Secretaría Distrital de Salud (SDS) a Capital Salud, por lo que es importante utilizar esta herramienta para simplificar los procesos administrativos (autorizaciones, facturación y auditoría). Lo único pendiente es la orden gerencial para que Capital Salud acceda a la historia clínica y esta regla sea una realidad.

2.3. Aumento en la calidad y el impacto en salud

En este NMR, la calidad determina el pago total, pues su resultado es un factor que condiciona la tarifa; así, para recibir el valor pactado es necesario cumplir los indicadores mínimos de calidad: en caso de no hacerlo, la tarifa disminuye proporcionalmente hasta el porcentaje acordado. El NMR mide el cumplimiento de la calidad a través de una batería de indicadores (inicialmente, doce); cada uno de ellos, en una línea de base observada, una meta y un mínimo aceptable.

Por otra parte, el nuevo Modelo de Atención Integral en Salud (MIAS) busca mayor prevención y menor atención, traducido ello en menor carga asistencial y menor gasto en salud en el Distrito Capital, premisa que opera en el NMR al reconocer las actividades de

protección específica y detección temprana (PyD) por evento, y no por capitación, para estimular su ejecución, y así permitir que los ahorros que tenga el asegurador en la carga asistencial se traduzcan en el aumento de las tarifas de las actividades PyD en futuros contratos por suscribir entre las partes.

2.4. ¿Cómo opera el mecanismo de pago por desempeño en el modelo PGAFID?

El modelo opera a partir de la facturación que realiza cada subred a Capital Salud por las actividades finales efectivamente dadas a la población. El modelo se basa en la estimación de una tarifa individual para cada actividad, para cada subred en cada periodo. El cálculo de dicha tarifa será determinado por tres componentes: 1) un valor de referencia, 2) un componente de producción y 3) un componente de calidad. A continuación, se presenta todo el modelo:

2.5. Factura total

La factura total que cada subred j cobrará a Capital Salud en el periodo t (F_{jt}) estará definida por la siguiente expresión:

$$F_{jt} = \sum_i^{30} T_{ijt} P_{ijt}$$

(1)

Donde:

F_{jt} : Factura total de la subred j en el periodo t

T_{ijt} : Tarifa de la actividad final i para la subred j en el periodo t

P_{ijt} : Producción de la actividad final i en el periodo t por la subred j

Otras formas de ver la ecuación (1) son:

$$F_{jt} = T_{1jt}P_{1jt} + T_{2jt}P_{2jt} + \dots + T_{30jt}P_{30jt} = \sum_i^{30} T_{ijt} P_{ijt}$$

$$F_{jt} = F_{1jt} + F_{2jt} + \dots + F_{30jt} = \sum_i^{30} F_{ijt}$$

Por otro lado, la tarifa de la actividad final i para la subred j en el periodo t (T_{ijt}), está definida por la siguiente expresión:

$$T_{ijt} = T_i^* * (\beta_{jt} + \alpha_{ijt})$$

(2)

Donde:

T_{ijt} : Tarifa de una actividad final i para una subred j , con $i \in \{1,2, \dots, 30\}$ y $j \in \{1,2,3,4\}$

T_i^* : Valor de referencia de una actividad final i .¹

β_{jt} : Componente de calidad de la tarifa para una subred j . $\beta_j \in [0.9,1.0]$

α_{ijt} : Componente de producción de la tarifa de una actividad final i para una subred j .

$\alpha_{ij} \in [0.0, \varphi_{ij}]$, donde, φ_{ij} es el límite superior del componente de producción (más adelante se explica con detalle).

3. Condiciones para implementar el NMRAFID

Existen unas condiciones para implementar el nuevo modelo. Cada una de estas es *muy importante* para el modelo, y, por lo tanto, deben ser definidas con mucha precaución. Estas condiciones son:

¹ Este valor de referencia fue obtenido como el costo promedio histórico de la actividad final más cercano al valor de eficiencia, realizado por las subredes. Está presentado en el documento *Marco Lógico*.

1. Se define un rango *acordado* (senda) entre las subredes y Capital Salud, con una producción mínima (P_{ijt}^{min}) y máxima (P_{ijt}^{max}), para cada actividad final i para la subred j en el periodo t . Se debe cumplir que:

$$P_{ijt-1} \ll P_{ijt}^{min} < P_{ijt}^{max}$$

Donde: P_{ijt-1} es la producción de la actividad final i para la subred j en el periodo anterior² $t - 1$.

Es decir, se espera que la diferencia entre P_{ijt}^{min} y P_{ijt-1} sea lo suficientemente amplia para que existan incentivos para aumentar la producción. Adicionalmente, se espera que la diferencia entre P_{ijt}^{max} y P_{ijt}^{min} no sea tan amplia.

Estos dos valores deben ser *acordados* entre las subredes y Capital Salud. P_{ijt}^{min} debe corresponder a la *mínima producción aceptable* que se espera obtener de la actividad final i , para entregar todo el *presupuesto global prospectivo* (PGP).

2. Se debe definir un PGP para cada subred j y para cada periodo t (PGP_{jt}), que será *acordado ex ante* entre Capital Salud y cada una de las subredes. Este PGP_{jt} estará relacionado con los valores históricos pagados, con la proporción de UPC disponible y con las actividades finales esperadas³.
3. P_{ijt}^{max} debe corresponder a la máxima *producción* que se puede obtener de la actividad final i al valor de referencia T_i^* con el PGP, es decir: el PGP del contrato debe estar condicionado por:

² Es importante calificar la producción del periodo anterior, ya que este es el punto de partida para determinar la eficiencia, y con ella, los estímulos.

³ Se sugiere que el PGP_{jt} sea el valor histórico pagado por Capital Salud a cada una de las subredes. Lo que el modelo pretende es que este valor no se modifique dramáticamente, pero que las subredes aumenten la producción, de acuerdo con las necesidades de la población de Capital y con una calidad mínima aceptable.

$$PGP_{jt} = \sum_i^{30} T_i^* * P_{ijt}^{max}$$

Nota: esto quiere decir que el valor facturado para cada actividad final i por cada subred j en el periodo t será, máximo, el producto del valor de referencia por la producción máxima esperada P_{ijt}^{max} . Ello implica directamente que el valor facturado total nunca excederá el PGP.

En caso de que la producción del periodo exceda la máxima, las partes deben analizar si es necesario acordar más producción aumentando el PGP pagado a valor de referencia.

3.1. ¿Cómo funciona cada componente del modelo?

Para el funcionamiento del modelo, cada actividad se debe evaluar por separado, pues (como ya se indicó) el modelo determina la tarifa por pagar para cada actividad final i por cada subred j en el periodo t . Para ello, utilizaremos el rango máximo pactado discriminado por cada actividad final i expresado, así⁴:

$$PGP_{jt} = \sum_i^{30} PGP_{ijt} = \sum_i^{30} T_i^* * P_{ijt}^{max}$$

Esta tarifa (como se muestra en la ecuación (2)) tiene tres componentes: 1) un valor de referencia, 2) un componente de producción y 3) un componente de calidad. A continuación, se definen los últimos dos componentes.

⁴ Las expresiones PGP_{ijt} y F_{ijt} solo se utilizan para explicar en el mecanismo de cálculo que el valor pactado por cada actividad no excede la producción máxima por el valor de referencia, pero no implican la existencia en el modelo de remuneración PGPAFED de 30 PGP ni 30 facturas.

3.2. Componente de producción: α_{ijt}

Para determinar el valor del componente de producción —y, por consiguiente, el de la factura (asociado a este componente) —, se definen tres casos, descritos a continuación:

3.2.1. Caso 1. Producción dentro del rango esperado o mayor a este (producción mayor que la P_{ijt}^{min})

Si la producción de la actividad final i por cada subred j en el periodo t supera el valor mínimo del rango acordado, entonces la factura será igual al valor asociado a esta actividad final, y acordado en el PGP multiplicado por el componente de calidad. Es decir,

$$P_{ijt}^{min} \leq P_{ijt} \rightarrow F_{ijt} = PGP_{ijt} * \beta_{jt}$$

Racionalidad: en un escenario ideal, se espera que la subred alcance el rango esperado; es decir, que su producción observada sea mayor que la mínima y menor que la máxima acordada. Si esto ocurre, la subred recibe el monto acordado para esa actividad⁵. Si el PGP acordado para dicha actividad es el valor histórico que ha pagado Capital Salud, entonces, Capital va a pagar lo mismo por recibir más producción.

Adicionalmente, si la subred produce por encima de la producción máxima acordada, igualmente recibirá el valor del PGP para esa actividad; y como $PGP_{ijt} = T_i^* * P_{ijt}^{max}$, entonces por cada unidad adicional que produzca la subred por encima del máximo, la tarifa será más baja que el valor de referencia. Por lo tanto, no existirá incentivo para que se produzca por encima de este valor.

⁵ Asumiendo que se cumple con el componente de calidad; es decir, $\beta_{jt} = 1$.

3.2.2. Caso 2. Producción inferior a la producción en el periodo anterior (producción menor que en el periodo $t - 1$)

Si la producción de la actividad final i por cada subred j en el periodo t es menor que la producción de esa actividad final en el periodo anterior de referencia, entonces el valor de la tarifa será igual al valor de referencia por el componente de calidad.

$$P_{ijt} < P_{ijt-1} \rightarrow \alpha_{ijt} = 0$$

Lo cual implica que

$$T_{ijt} = T_i^* * \beta_{jt} \rightarrow F_{ijt} = T_i^* * \beta_{jt} * P_{ijt}$$

Racionalidad: si la producción de la subred está por debajo de la producción observada en el periodo anterior, quiere decir que la subred empeoró en su producción respecto al periodo de observación anterior. Al utilizar valor de referencia como tarifa (suponiendo que se cumple calidad), la subred no alcanzaría a asumir los costos de producción. No existiría un incentivo a estar por debajo de este valor.

3.2.3. Caso 3. Producción mayor que el periodo anterior y menor que el rango esperado (producción mayor o igual a la producción en el periodo $t - 1$ y menor o igual a la mínima acordada).

En este caso, si la producción de la actividad final i por cada subred j en el periodo t supera el valor del periodo anterior, pero no alcanza a llegar al rango, entonces la tarifa aumentará proporcionalmente al grado de avance respecto a estos puntos de referencia, sin que supere el PGP de la actividad. La siguiente expresión define esto:

$$P_{ijt-1} \leq P_{ijt} \leq P_{ijt}^{min} \rightarrow \alpha_{ijt} = \frac{P_{ijt} - P_{ijt-1}}{P_{ijt}^{min} - P_{ijt-1}} * \varphi_{ij}$$

3.2.4. ¿Qué es φ_{ij} ?

Como ya se dijo, φ_{ij} es el valor máximo del componente de producción de la tarifa. Este valor máximo debe ajustar el valor máximo del PGP para la actividad. Por lo tanto,

$$\varphi_{ij} = \frac{P_{ijt}^{max}}{P_{ijt}^{min}} - 1$$

El valor de φ_{ij} siempre será mayor que cero, debido a que $P_{ijt}^{min} < P_{ijt}^{max}$. Y para garantizar que $\varphi_{ij} < 1$, debemos asegurar que: $0.5P_{ijt}^{max} < P_{ijt}^{min} < P_{ijt}^{max}$. El anexo 1 muestra de dónde se obtiene φ_{ij} .

Racionalidad: si la subred avanza en relación con el periodo anterior, pero no alcanza el rango esperado, la tarifa va a aumentar proporcionalmente al avance, sin que exceda el PGP.

4. Componente de calidad: β_j

El desempeño clínico estará representado por indicadores clínicos de calidad en la atención. Estos indicadores deberán ser concertados entre las dos partes: para el caso, las subredes y EPS Capital Salud. Como ya se mencionó, se espera que las subredes cumplan con un *mínimo* de estándares de calidad. Esto justifica que las subredes puedan ser castigadas hasta con el 10 % en la tarifa si no se cumplen con los estándares mínimos de calidad. Como se observa en el subíndice de β_j , este factor se calculará para toda la subred j y afectará todas las actividades finales i por igual. La siguiente expresión representa a β_j :

$$\beta_j = 0.9 + 0.1[\min\{I_{1j}, I_{2j}, \dots, I_{Kj}\}]$$

Donde: K representa el número de indicadores k que deberán ser concertados entre las dos partes. Se sugiere que K no sea muy grande: esto da simplicidad y fácil comprensión al mecanismo.

Por otra parte, la definición de cada indicador I_{kj} , sus características, su forma de medición y sus fuentes de información para su cálculo deben ser definidas y explícitas. Cada una de ellas debe ser definida siguiendo las recomendaciones internacionales o nacionales al respecto, y con la participación y la conformidad de las partes. Adicionalmente, $I_{kj} \in [0,1]$; es decir, cada indicador debe representar a una proporción de cumplimiento del indicador k .

Nota 1: es importante hacer énfasis en que las sendas (rango entre producción mínima y máxima esperada) deben ser estimadas de la forma más precisa posible. Estas sendas son el resultado de un acuerdo entre Capital Salud y las subredes, donde la primera expone las necesidades totales de su población asegurada (demanda) y las segundas presentan su capacidad máxima (oferta). Para ello, es muy importante clarificar y utilizar como referencia la producción del periodo anterior. Se espera que el mecanismo propenda por que las subredes vayan adaptando su oferta a las necesidades de la población de Capital Salud, de tal forma que en el futuro las subredes atiendan toda la demanda de la población de Capital Salud, sin un detrimento en la calidad de la atención y sin exceder el 92 % de la UPC.

Nota 2: considerando que no se ha pactado entre las partes la forma de medición del β_j y que a la fecha se va a terminar el II trimestre de ejecución del modelo, se sugiere que para facturar estos dos trimestres transcurridos, el β_j sea igual a 1 y se calcule a partir del III trimestre.

Ejemplo⁶:

⁶ Para analizar por separado el comportamiento de la eficiencia administrativa o el componente de producción (α_{ij}), asumiremos la calidad β_j como 1, para que no tenga efecto matemático.

La subred j vendió a Capital Salud en el periodo anterior ($t - 1$) un total de 90 actividades de la tipología i por un valor total de \$1000; y, conociendo que, según el NMR, la tarifa eficiente de la actividad i (T_i^*) es: \$8, la subred j pactó para este periodo un valor total de \$1000 por una producción de actividades i entre 100 y 125.

Donde:

$$P_{ij}^{min} = 100$$

$$P_{ij}^{max} = 125$$

$$P_{ijt-1} = 90$$

$$T_i^* = \$8$$

$$PGP_{ijt} = \$1000$$

F_{ijt} : valor de la factura total de la subred j en el periodo t correspondiente a la actividad final i

Analicemos los tres posibles casos:

4.1. Caso 1: Producción dentro del rango esperado o mayor que este (producción mayor que la P_{ijt}^{min})

$$P_{ijt} = 113 \rightarrow P_{ijt}^{min} \leq P_{ijt} \rightarrow F_{ijt} = PGP_{ijt} * \beta_{jt}$$

$$F_{ijt} = PGP_{ijt} * \beta_{jt}$$

$$F_{ijt} = \$ 1.000 * \beta_{jt}$$

$$F_{ijt} = \$ 1.000 * 1$$

$$F_{ijt} = \$1.000$$

4.2. Caso 2. Producción inferior a la producción en el periodo anterior (producción menor que en el periodo $t - 1$)

$$P_{ijt} = 89 \rightarrow P_{ijt} < P_{ijt-1} \rightarrow \alpha_{ijt} = 0$$

Lo cual implica que:

$$T_{ijt} = T_i^* * \beta_{jt} \rightarrow F_{ijt} = T_i^* * \beta_{jt} * P_{ijt}$$

$$F_{ijt} = 8 * 1 * 89$$

$$F_{ijt} = 712$$

4.3. Caso 3. Producción mayor que el periodo anterior y menor que el rango esperado (producción mayor o igual que la producción en el periodo $t - 1$ y menor o igual que la mínima acordada)

$$P_{ijt} = 95 \rightarrow P_{ijt-1} \leq P_{ijt} \leq P_{ijt}^{min} \rightarrow \alpha_{ijt} = \frac{P_{ijt} - P_{ijt-1}}{P_{ijt}^{min} - P_{ijt-1}} * \varphi_{ij}$$

Hallamos φ_{ij}

$$\varphi_{ij} = \frac{P_{ijt}^{max}}{P_{ijt}^{min}} - 1$$

$$\varphi_{ij} = \frac{125}{100} - 1$$

$$\varphi_{ij} = 1.25 - 1$$

$$\varphi_{ij} = 0.25$$

Hallamos α_{ijt}

$$\alpha_{ijt} = \frac{P_{ijt} - P_{ijt-1}}{P_{ijt}^{min} - P_{ijt-1}} * \varphi_{ij}$$

$$\alpha_{ijt} = \frac{95 - 90}{100 - 90} * \varphi_{ij}$$

$$\alpha_{ijt} = \frac{5}{10} * \varphi_{ij}$$

$$\alpha_{ijt} = 0.5 * \varphi_{ij}$$

Remplazamos φ_{ij}

$$\alpha_{ijt} = 0.5 * \varphi_{ij}$$

$$\alpha_{ijt} = 0.5 * 0.25$$

$$\alpha_{ijt} = 0.125$$

Remplazamos α_{ijt}

$$T_{ijt} = T_i^* * (\beta_{jt} + \alpha_{ijt})$$

$$T_{ijt} = 8 * (1 + 0.125)$$

$$T_{ijt} = 8 * (1.125)$$

$$T_{ijt} = \$9$$

Hallamos F_{ijt}

$$F_{ijt} = T_{ijt} P_{ijt}$$

$$F_{ijt} = 9 * 95$$

$$F_{ijt} = \$855$$

Referencias

Cashin, C., Chi, Y-L., Smith, P. C., Borowitz, M., & Thomson, S. (2014). *Paying for performance in health care: Implications for health system performance and accountability*. European Observatory on Health Systems and Policies Series. Open Univ. Press.

Miller, G., & Babiarz, K. S. (2014). Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs. En: *Encyclopedia of Health Economics* (pp. 457-466). Elsevier.

Elaboró:

Equipo Técnico NMR PGPAFED (SDS)

Asesores Uniandes

Anexo 1. Identificación de φ_{ij}

Por definición, tenemos que

$$PGP_{ijt} = T_i^* * P_{ijt}^{max} \quad (3)$$

Esto significa que para el cálculo del PGP, se asume que el valor de $\beta_{jt} = 1$ y de $\alpha_{ijt} = 0$.

También sabemos, por el caso 1, que si $P_{ijt}^{min} \leq P_{ijt}$, entonces, $F_{ijt} = PGP_{ijt} * \beta_{jt}$.
Suponiendo que la calidad sea la esperada —es decir $\beta_{jt} = 1$ —, entonces la siguiente expresión es correcta:

$$PGP_{ijt} = (1 + \varphi_{ij}) * T_i^* * P_{ijt}^{min} \quad (4)$$

Igualando (3) y (4) y despejando φ_{ij} , tenemos:

$$(1 + \varphi_{ij}) * T_i^* * P_{ijt}^{min} = T_i^* * P_{ijt}^{max}$$

$$(1 + \varphi_{ij}) * P_{ijt}^{min} = P_{ijt}^{max}$$

$$1 + \varphi_{ij} = \frac{P_{ijt}^{max}}{P_{ijt}^{min}}$$

$$\varphi_{ij} = \frac{P_{ijt}^{max}}{P_{ijt}^{min}} - 1$$