

Artículo: Original

Efficiency of four probability sampling designs: a Monte Carlo simulation study

Eficiencia de cuatro diseños de muestreo probabilístico: estudio de simulación Monte Carlo

¹Marco Antonio Ramos Espinal  .

¹Universidad Nacional Autónoma de Honduras (UNAH), Ciudad Universitaria, Tegucigalpa, Honduras

***Autor de correspondencia:** Marco Antonio Ramos Espinal 

Enviado: 13-07-2026

Revisado: 19-08-2026

Aceptado: 03-09-2026

Publicado: 12-10-2026

Como citar:

Ramos Espinal, M. A. (2026). Eficiencia de cuatro diseños de muestreo probabilístico: estudio de simulación Monte Carlo. *Sapiens International Multidisciplinary Journal*, 3(5), 1-23. <https://doi.org/10.71068/SIMJ0005>.

ABSTRACT

Introduction: the choice of sampling design determines the precision and validity of survey inferences, yet differences between designs are rarely quantified on a population with fully known parameters. We compared four probability sampling designs (simple random, stratified, cluster and systematic) for estimating a mean, a standard deviation and a proportion.

Method: under specified assumptions, we generated a synthetic population of $N = 1,000,000$ individuals (age, weight and sex) in 40,000 clusters of 25. For each design we drew 1,000 samples of $n = 2,000$ and computed estimates and 90%, 95% and 99% confidence intervals. We assessed bias, root mean squared error, relative efficiency, design effect, coverage and interval width, with a sensitivity analysis of standard errors.

Results: all designs were practically unbiased (maximum relative bias 0.197%). Age stratification gained only for age, the stratification variable (relative efficiency 10.37 for the mean and 2.94 for the standard deviation), not for weight or the proportion of men; its mean relative efficiency (3.29) falls to 1.52 without mean age. Systematic sampling on a frame ordered by an age-related variable gained moderately (2.01 for mean age). Approximate standard errors produced conservative intervals (99.9% and 99.1% coverage at 95%); design-based standard errors brought them to 93.6% and 95.4%. Cluster sampling was the least efficient (design effects up to 2.85).

Conclusions: no design dominated; efficiency depended on how the design relates to the parameter. Because all designs used the same sample size but were not evaluated at equal operational cost, the comparison concerns statistical efficiency, not economic superiority. As the results come from one synthetic population with age-defined strata, the magnitudes should be generalized cautiously.

Keywords: probability sampling; stratified sampling; cluster sampling; systematic sampling; Monte Carlo simulation; design effect

RESUMEN

Introducción: el diseño de muestreo determina la precisión de las inferencias, pero sus diferencias rara vez se cuantifican con parámetros conocidos. Comparamos cuatro diseños probabilísticos (aleatorio simple, estratificado, por conglomerados y sistemático) para estimar una media, una desviación típica y una proporción.

Método: bajo supuestos especificados, generamos una población sintética de $N = 1\,000\,000$ de personas (edad, peso y sexo) en 40 000 racimos de 25. Por diseño extrajimos 1 000 muestras de $n = 2\,000$ con estimaciones e intervalos de confianza al 90 %, 95 % y 99 %. Evaluamos sesgo, error cuadrático medio, eficiencia relativa,

efecto de diseño, cobertura y amplitud de intervalos.

Resultados: los cuatro diseños fueron prácticamente insesgados (sesgo relativo $\leq 0,197\%$). Estratificar por edad ganó solo en la edad, variable de estratificación (eficiencia relativa: 10,37 en la media; 2,94 en la desviación típica), no en el peso ni en la proporción de hombres; su eficiencia media (3,29) baja a 1,52 sin la media de la edad. El sistemático con marco ordenado por una variable asociada a la edad ganó moderadamente (2,01). Los errores estándar aproximados dieron coberturas excesivas (99,9 % y 99,1 % al 95 %), corregidas a 93,6 % y 95,4 % con errores coherentes con el diseño. Los conglomerados fueron los menos eficientes (efecto de diseño $\leq 2,85$).

Conclusiones: ningún diseño dominó; la eficiencia dependió de la relación entre el diseño y el parámetro. Como los diseños compartieron tamaño muestral, pero no se compararon a igual costo operativo, los resultados se refieren a eficiencia estadística, no a superioridad económica. La ventaja del estratificado depende de los estratos definidos por edad y de los supuestos de esta población sintética.

Palabras clave: muestreo probabilístico; muestreo estratificado; muestreo por conglomerados; muestreo sistemático; simulación Monte Carlo; efecto de diseño

INTRODUCCIÓN

El muestreo probabilístico sostiene la inferencia sobre poblaciones finitas desde que Neyman formalizó en 1934 el método representativo (1), y el estimador de Horvitz y Thompson (2) completó la base. Si se conoce la probabilidad con que cada unidad entra en la muestra, se puede estimar el parámetro y también la precisión de esa estimación sin hacer supuestos sobre la forma de la población (3-6). Esa propiedad explica que siga siendo la referencia frente a las muestras no probabilísticas, cada vez más frecuentes (7,8).

Quien diseña una encuesta suele tener al menos cuatro opciones clásicas: el muestreo aleatorio simple (MAS), el estratificado, el muestreo por conglomerados o racimos y el sistemático (5,9). Cada uno organiza la población de un modo distinto. El estratificado aprovecha subpoblaciones homogéneas definidas de antemano; el de conglomerados ahorra desplazamientos al agrupar unidades vecinas; el sistemático recorre un marco ordenado a intervalos fijos y con frecuencia introduce una estratificación implícita (10-12). Elegir uno u otro cambia la varianza de los estimadores, la amplitud de los intervalos y, por tanto, el costo de alcanzar una precisión determinada (13,14).

La teoría de estas diferencias está madura, pero su magnitud concreta rara vez se observa en la práctica, porque en una encuesta real el parámetro verdadero es desconocido y el error de estimación no puede medirse. La simulación Monte Carlo sobre una población artificial conocida en su totalidad resuelve el problema (15,16). Con la media, la desviación típica y la proporción reales a la vista, el sesgo, el error cuadrático medio y la cobertura de los intervalos se calculan directamente y pueden compararse entre diseños (17). Las poblaciones sintéticas y los «gemelos digitales» se usan cada vez más con este fin, por ejemplo para evaluar diseños de inventario forestal (18).

Las comparaciones disponibles suelen ocuparse de un diseño o de una familia de estimadores a la vez: nuevos estimadores de la media o de la varianza bajo un diseño fijo (19-22), variantes del sistemático para marcos con tendencia lineal (12), estimadores de la varianza del sistemático (23) o reglas de asignación óptima en el estratificado (24,25). Es menos frecuente una comparación

simultánea de los cuatro diseños básicos sobre una misma población conocida y con un mismo tamaño de muestra, que evalúe a la vez la estimación puntual y la validez de los intervalos para parámetros de distinta naturaleza: una media, una desviación típica y una proporción. Ese es el vacío que aborda este estudio. Además de ordenar los diseños por eficiencia, mostramos para qué parámetros se produce la ganancia de cada diseño, cuándo esa ganancia no se traslada a los intervalos de confianza porque el error estándar empleado no la recoge, y cuánto cambia la cobertura al usar un error estándar coherente con el diseño.

En este trabajo simulamos una población de 1 000 000 de personas con edad, peso y sexo, y comparamos los cuatro diseños en la estimación de una media, una desviación típica y una proporción. El objetivo general fue determinar qué diseño resulta más eficiente y en qué condiciones. Los objetivos específicos fueron (i) cuantificar el sesgo y la precisión de cada estimador puntual, (ii) evaluar la cobertura empírica y la amplitud de los intervalos de confianza al 90 %, 95 % y 99 %, (iii) ordenar los diseños según su desempeño global y (iv) valorar la sensibilidad de la cobertura al método de cálculo del error estándar. Partimos de la hipótesis, derivada de la teoría, de que ningún diseño domina en todos los parámetros y de que la eficiencia depende de la relación entre la estructura del diseño y lo que se estima (3,4).

El aporte tiene una vertiente metodológica y otra aplicada. El estudio sigue el esquema ADEMP para simulaciones (15,17) y atiende las recomendaciones de documentar los procedimientos para favorecer su replicación (26) y de evitar comparaciones que favorezcan de antemano a algún método (27). Se acompañan scripts de comprobación independientes del código de producción; su correspondencia exacta con los resultados publicados debe contrastarse antes de la entrega editorial. Se ofrecen además orientaciones para planificar encuestas comunitarias y de salud poblacional, donde la edad, el peso y el sexo son variables de uso corriente.

REVISIÓN DE LA LITERATURA

Consideramos una población finita $U = \{1, 2, \dots, N\}$ en la que cada unidad i tiene un valor y_i de la variable de interés y , para un atributo dicotómico, un indicador $a_i \in \{0, 1\}$. Los parámetros objetivo son la media μ , la desviación típica σ y la proporción P de la ecuación (1) (2,3).

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i, \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2}, \quad P = \frac{1}{N} \sum_{i=1}^N a_i \quad (1)$$

Un diseño de muestreo es una distribución de probabilidad sobre el conjunto de muestras posibles, que asigna a cada unidad una probabilidad de inclusión π_i conocida y positiva. La fracción de muestreo es $f = n/N$ y el factor de corrección por población finita, $1 - f$ (4).

2.1 Muestreo aleatorio simple

En el MAS sin reposición todas las muestras de tamaño n tienen la misma probabilidad y $\pi_i = n/N$. La media muestral de la ecuación (2) es insesgada para μ (3,5).

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2)$$

Las ecuaciones (3) y (4) dan su varianza y la del estimador de la proporción p , con S^2 como cuasivarianza poblacional. Como no usa información auxiliar ni estructura alguna de la población, el MAS sirve de referencia para medir la eficiencia de los demás diseños (5).

$$\text{Var}(\bar{y}) = (1 - f) \frac{S^2}{n} \quad (3)$$

$$\text{Var}(p) = (1 - f) \frac{P(1-P)}{n-1} \quad (4)$$

2.2 Muestreo aleatorio estratificado

El diseño estratificado divide la población en L estratos de tamaño N_h y extrae de cada uno una submuestra independiente por MAS. El estimador de la media es la media ponderada de la ecuación (5) y su varianza, la de la ecuación (6) (1,3).

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h, \quad W_h = \frac{N_h}{N} \quad (5)$$

$$\text{Var}(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{S_h^2}{n_h} \quad (6)$$

La varianza depende solo de la variabilidad dentro de los estratos, de modo que cuanto más homogéneos sean respecto de la variable de interés, menor será (4). Con asignación proporcional ($n_h = n \cdot W_h$) el diseño nunca es menos eficiente que el MAS para la media, y gana más cuando la variable de estratificación está muy asociada con la de estudio (13,14). La asignación óptima de Neyman afina todavía esa ganancia (1,24).

2.3 Muestreo por conglomerados (racimos)

En este diseño la población se organiza en G racimos, se seleccionan g de ellos y se observan todas sus unidades (muestreo de una etapa). Para racimos de igual tamaño M elegidos por MAS, el estimador de la media y su varianza, que depende de la variabilidad entre racimos S_b^2 , son los de las ecuaciones (7) y (8) (3,13).

$$\bar{y}_{cl} = \frac{1}{g} \sum_{i=1}^g \bar{y}_i \quad (7)$$

$$\text{Var}(\bar{y}_{cl}) = \left(1 - \frac{g}{G}\right) \frac{S_b^2}{g} \quad (8)$$

Kish resumió su eficiencia en el efecto de diseño, que para racimos de igual tamaño se aproxima con la ecuación (9), donde ρ es el coeficiente de correlación intraclase (13).

$$\text{Deff} \approx 1 + (M - 1)\rho \quad (9)$$

Las unidades de un mismo racimo suelen parecerse ($\rho > 0$) y por eso, con igual número de unidades, el diseño tiende a ser menos preciso que el MAS. Se elige por el ahorro en trabajo de campo y porque a menudo solo existen marcos de racimos (13,28).

2.4 Muestreo sistemático

Sobre un marco ordenado de N unidades se elige un arranque aleatorio $r \in \{1, \dots, k\}$ y luego cada k -ésima unidad, con el intervalo de muestreo de la ecuación (10).

$$k = \frac{N}{n} \quad (10)$$

El resultado depende del orden del marco (10,11). Si el marco sigue una variable asociada con la de interés, el diseño se comporta como un estratificado implícito y reduce la varianza (11,12); si el orden es aleatorio respecto de esa variable, equivale al MAS. Con una sola muestra sistemática (un único arranque aleatorio) no existe un estimador insesgado de la varianza. En la práctica se usa la fórmula del MAS, que la sobreestima cuando el marco tiene tendencia, o estimadores aproximados basados en las diferencias entre unidades contiguas del marco (23,29), como el de la ecuación (14) de la sección 2.5.

2.5 Estimación puntual y por intervalos

La estimación puntual entrega un valor $\hat{\theta}$ para el parámetro θ . El intervalo de confianza de nivel $1 - \alpha$ de la ecuación (11) añade la incertidumbre, con $EE(\hat{\theta})$ como error estándar estimado y $z_{1-\alpha/2}$ como cuantil normal (1,645, 1,960 y 2,576 para el 90 %, 95 % y 99 %).

$$\hat{\theta} \pm z_{1-\alpha/2} EE(\hat{\theta}) \quad (11)$$

Para medias y proporciones, el error estándar sale de las fórmulas de varianza de cada diseño (ecuaciones 3 a 8). En la proporción empleamos el intervalo de Wald, aunque los de Wilson (30) y de Agresti y Coull (31) cubren mejor en muestras pequeñas o con proporciones extremas (32,33). Para la desviación típica, el análisis principal usa la aproximación normal de la ecuación (12), válida con n grande (29). Esta fórmula no depende del diseño: supone una muestra aleatoria simple de una variable con la curtosis de la distribución normal, de modo que puede sobrestimar o subestimar el error cuando el diseño gana precisión o la variable no es normal (34,35).

$$EE(s) \approx \frac{s}{\sqrt{2(n-1)}} \quad (12)$$

Como alternativa coherente con el diseño usamos la linealización de Taylor (29). La desviación típica se aproxima por la media de la variable de influencia de la ecuación (13), y su error estándar se obtiene aplicando a esa media la fórmula de varianza del diseño: la del MAS (ecuación 3), la estratificada (ecuación 6) o la de racimos (ecuación 8). En el sistemático, en lugar de la fórmula del MAS, puede emplearse el estimador de diferencias sucesivas de la ecuación (14), que compara cada unidad con la anterior en el orden del marco y recoge así parte de la estratificación implícita (23,29).

$$u_i = \frac{(y_i - \bar{y})^2 - s^2}{2s} \quad (13)$$

$$v_{DS}(\bar{y}) = \frac{1-f}{2n(n-1)} \sum_{i=2}^n (y_i - y_{i-1})^2 \quad (14)$$

2.6 Medidas de eficiencia y bondad de ajuste

Como θ es conocido, el desempeño de cada estimador se mide sobre las B réplicas de la simulación (15,17). La ecuación (15) define el sesgo y la (16), el error cuadrático medio (ECM) y su raíz (RECM), que expresa el error total en las unidades del parámetro.

$$\text{Sesgo}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b - \theta \quad (15)$$

$$\text{ECM} = \text{Sesgo}^2 + \text{Var}(\hat{\theta}), \quad \text{RECM} = \sqrt{\text{ECM}} \quad (16)$$

La eficiencia relativa frente al MAS (ecuación 17) compara errores cuadráticos medios, y el efecto de diseño (ecuación 18) compara varianzas. Una eficiencia mayor que 1 o un Deff menor que 1 indican ventaja sobre el MAS (13).

$$\text{Ef}_{\text{rel}} = \frac{\text{ECM}_{\text{MAS}}}{\text{ECM}} \quad (17)$$

$$\text{Deff} = \frac{\text{Var}_{\text{dis}}(\hat{\theta})}{\text{Var}_{\text{MAS}}(\hat{\theta})} \quad (18)$$

En los intervalos, la cobertura empírica de la ecuación (19) es la proporción de réplicas en que el intervalo contiene a θ , idealmente igual al nivel nominal, y la amplitud media mide su precisión. Con igual cobertura preferimos el intervalo más estrecho (5,17).

$$\hat{C} = \frac{1}{B} \sum_{b=1}^B \mathbf{1}\{\theta \in \text{IC}_b\} \quad (19)$$

MÉTODO

Diseñamos un experimento de simulación Monte Carlo en el que la población se conoce por completo y los cuatro diseños se aplican repetidamente en condiciones idénticas. Seguimos el esquema de Morris et al. (15), conocido como ADEMP y desarrollado en plantillas recientes para planificar y comunicar simulaciones (17): objetivos, mecanismo generador de datos, estimandos, métodos y medidas de desempeño.

3.1 Población sintética

Generamos $N = 1\,000\,000$ de individuos agrupados en $G = 40\,000$ racimos de $M = 25$ unidades, que representan agregados territoriales como manzanas o sectores censales. Para imitar el parecido entre vecinos añadimos efectos aleatorios de racimo normales a la edad y al peso (desviación típica de 3 años y de 3 kg), lo que produce correlación intraclase positiva. El sexo es una variable de Bernoulli cuya probabilidad de ser hombre es 0,49 más un efecto normal de racimo con desviación típica de 0,05 (acotada entre 0,25 y 0,75). La edad combina una media de 42 años, el efecto del racimo y un término individual normal con desviación típica de 14 años, y se truncó al intervalo [18, 95]; el truncamiento acumula al 4,7 % de la población en 18 años. El peso es 62 kg más 13 kg en los hombres, 0,30 kg por cada año de diferencia respecto de los 42 años, el efecto del racimo y un término individual con desviación típica de 9 kg, y se truncó a [40, 140] kg. Con la población completa calculamos los parámetros verdaderos que sirven de patrón (Tabla 1).

Los parámetros se eligieron para representar la población adulta de una encuesta comunitaria o de salud, no para reproducir un país concreto. La media de 42 años y la desviación típica individual de

14 años, con límites de 18 y 95 años, dan una población de 18 años o más con una estructura etaria amplia, y el truncamiento evita valores fuera del dominio de una encuesta de adultos. La diferencia de 13 kg entre sexos y la pendiente de 0,30 kg por año generan asociaciones moderadas entre peso, sexo y edad (en la población, $r = 0,53$ entre sexo y peso y $r = 0,34$ entre edad y peso), mientras que la edad y el sexo se generaron de forma independiente ($r = 0,001$); el límite de 40 kg afecta solo al 0,8 % de la población. Los efectos de racimo producen correlaciones intraclases de 0,044 en la edad, 0,067 en el peso y 0,010 en el sexo, valores bajos o moderados como los que se registran en encuestas de hogares por conglomerados (13,28). Estas relaciones determinan qué parámetros pueden beneficiarse de la estratificación por edad y del orden del marco, y por eso las tratamos como condiciones del experimento y no como rasgos generales de cualquier población.

Tabla 1.

Parámetros del diseño de simulación y valores poblacionales verdaderos

Parámetro	Valor
Tamaño de la población (N)	1 000 000
Tamaño de muestra por réplica (n)	2 000 ($f = n/N = 0,2 \%$)
Número de muestras Monte Carlo (B)	1 000 por diseño
Estructura de racimos	40 000 racimos de tamaño $M = 25$
Intervalo de muestreo sistemático (k)	500
Estratos por grupos de edad (L)	4 ([18,30), [30,45), [45,60), [60,95])
Tamaños de estrato (Nh)	199 966; 381 477; 313 221; 105 336
Asignación proporcional (nh)	400; 763; 626; 211
Media real de la edad (μ)	42,34 años
Desviación típica real de la edad (σ)	13,75 años
Media real del peso (μ)	68,49 kg
Desviación típica real del peso (σ)	12,14 kg
Proporción real de hombres (P)	49,0 % (0,490)
Correlación de la variable auxiliar de ordenamiento (Z) con la edad	0,70 (realizada: 0,7002)
Semilla aleatoria (reproducibilidad)	20260627

Nota: elaboración propia a partir de la simulación (semilla 20260627).

3.2 Diseños de muestreo implementados

Aplicamos los cuatro diseños a la misma población con un tamaño de muestra común de $n = 2\,000$. (a) MAS: selección de 2 000 unidades sin reposición. (b) Estratificado: cuatro estratos por grupos de edad, una variable disponible en la mayoría de los marcos comunitarios, con asignación proporcional y MAS dentro de cada estrato. (c) Conglomerados: $g = 80$ racimos de 25 personas seleccionados por MAS, con observación de todas sus unidades. (d) Sistemático: el marco se ordenó de menor a mayor según una variable auxiliar $Z = 0,70 \cdot \text{edad} + (1 - 0,702)1/2 \cdot \epsilon$, donde edad es la edad estandarizada y $\epsilon \sim N(0, 1)$ un ruido independiente, de modo que $\text{corr}(Z, \text{edad}) = 0,70$ por construcción (0,7002 en la población generada). Z imita un registro administrativo asociado a la edad sin coincidir con ella. Se elige un arranque aleatorio r entre 1 y $k = 500$ y se toman las unidades $r, r + k, r + 2k, \dots$; como el diseño solo admite 500 muestras distintas, las réplicas eligen el arranque con reposición. En cada diseño usamos los estimadores y las fórmulas de varianza de la sección anterior. En el sistemático, el

análisis principal estima la varianza con la fórmula del MAS, como es habitual (29); esa fórmula es una aproximación y no un estimador insesgado, lo que tratamos como una limitación y examinamos en el análisis de sensibilidad. La igualdad de n controla el número de observaciones, pero no homogeneiza desplazamientos, elaboración del marco, personal, tiempo ni otros costos de ejecución; estos no se midieron ni se modelaron.

3.3 Procedimiento de simulación Monte Carlo

Para cada diseño extrajimos $B = 1\,000$ muestras independientes. En cada una calculamos cinco estimandos (media y desviación típica de la edad, media y desviación típica del peso y proporción de hombres), con sus errores estándar e intervalos al 90 %, 95 % y 99 %. Obtuvimos así, por diseño y estimando, una distribución muestral empírica de 1 000 valores. El análisis del manuscrito se describe como desarrollado en Python 3 con NumPy y una semilla fija (20260627). Para facilitar una comprobación independiente se adjunta el script de referencia S1 (Suplementario_S1_simulacion_muestreo.py), que implementa las decisiones metodológicas documentadas, y el script S2 (Suplementario_S2_figuras.py), que genera gráficos auxiliares a partir de la salida de S1. Estos archivos no son el código de producción original y no se afirma que reproduzcan exactamente los valores de las Tablas 1 a 6 o las Figuras 1 a 6.

Para cada estimando calculamos dos errores estándar. El principal (EE1) se usó en las Tablas 2 a 5: en las medias y la proporción, la fórmula de varianza de cada diseño (ecuaciones 3 a 8, con la del MAS en el sistemático); en la desviación típica, $s/[2(n - 1)]^{1/2}$ en los cuatro diseños (ecuación 12), donde s es la cuasidesviación muestral en el MAS, el sistemático y los racimos, y en el estratificado la raíz de $\Sigma W_h[sh^2 + (\bar{y}_h - \bar{y}_{st})^2]$. El alternativo (EE2) se usó en el análisis de sensibilidad (Tabla 6): en la desviación típica, la linealización de la ecuación (13) con la varianza del MAS, la estratificada (con u_i calculada respecto de $\bar{y}_{st}$ y de la desviación típica estratificada) o la de racimos (sobre las medias de u_i por racimo); en el sistemático, además, el estimador de diferencias sucesivas (ecuación 14) para las medias, la proporción y la variable de influencia. En los demás casos EE2 coincide con EE1. El cálculo de EE2 no consume números aleatorios, por lo que las estimaciones puntuales y EE1 son idénticos con o sin el análisis de sensibilidad.

3.4 Medidas de desempeño

Empleamos siete medidas: sesgo, error relativo porcentual, RECM, eficiencia relativa frente al MAS (ECMMAS/ECM), efecto de diseño (Deff), cobertura empírica a los tres niveles y amplitud media de los intervalos. Consideramos más eficiente al diseño que combina un ECM bajo con una cobertura próxima a la nominal. Para el balance global promediamos, en cada diseño, la eficiencia relativa de los cinco estimandos y la desviación absoluta de la cobertura respecto del 95 %. Como el promedio aritmético es sensible a valores extremos y depende de los estimandos elegidos, informamos también la mediana y el promedio sin la media de la edad. La expresión «eficiencia» se utiliza exclusivamente en su acepción estadística (comparación de errores de estimación); no se estimó la relación entre precisión, presupuesto ni costo de campo.

3.5 Análisis de los datos

El análisis tuvo la forma de un experimento factorial con tres factores: diseño (cuatro niveles), estimando (cinco) y nivel de confianza (tres). De ahí resultan 20 estimadores puntuales y 60 intervalos, cada uno evaluado sobre 1 000 réplicas. Los valores críticos normales quedaron fijados en el código y los errores estándar se calcularon como se describe en la sección 3.3. Como controles de validez comprobamos que los cuatro estimadores fueran prácticamente insesgados (el mayor sesgo relativo fue de 0,197 %, en la proporción de hombres bajo MAS) y que el MAS reprodujera coberturas cercanas a las nominales (15). También estimamos la correlación intraclase de la edad, el peso y el sexo con el estimador de análisis de varianza, para contrastar la ecuación (9) con los efectos de diseño observados. Al volver a ejecutar el código con la misma semilla obtuvimos valores idénticos a los que se reportan en las tablas. Las figuras se elaboraron con Matplotlib.

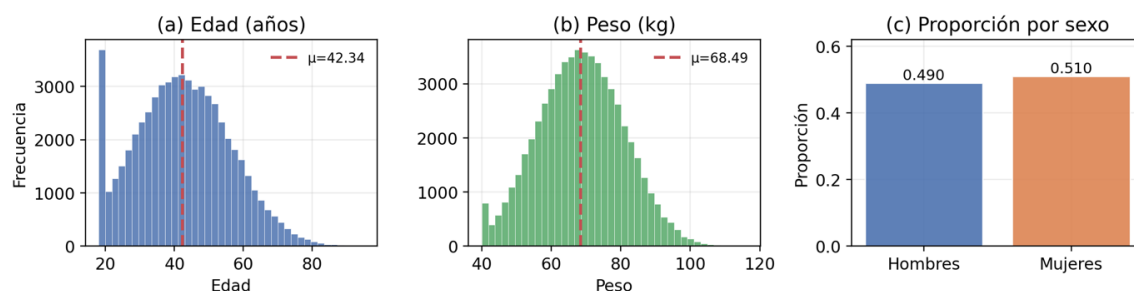
RESULTADOS

4.1 La población y los parámetros verdaderos

La población simulada tiene una edad media de 42,34 años, con desviación típica de 13,75 años; el peso medio es de 68,49 kg, con desviación típica de 12,14 kg, y el 49,0 % son hombres (Figura 1). La distribución del peso es casi simétrica y recoge la diferencia entre sexos incorporada en el modelo. La de la edad es unimodal, con cola hacia las edades mayores y una barra aislada en 18 años que corresponde al truncamiento.

Figura 1.

Distribución de la edad (a) y del peso (b) y proporción por sexo (c) en la población sintética (N = 1 000 000). La línea discontinua señala la media real



Nota: Elaboración propia a partir de la simulación (semilla 20260627).

4.2 Estimación puntual e intervalos en una muestra ilustrativa

La Tabla 2 recoge las estimaciones y los intervalos de una sola muestra por diseño para la media y la desviación típica del peso y para la proporción de hombres. Su propósito es solo ilustrar cómo se obtienen y se leen las estimaciones: el resultado de una muestra aislada depende del azar de esa extracción y no sirve para comparar diseños, por lo que las comparaciones de eficiencia se basan en las 1 000 réplicas resumidas en las Tablas 3 a 6. En esta muestra, los intervalos se ensanchan con el nivel de confianza, como corresponde, y los de conglomerados son los más anchos. Al 95 %, el de la media del peso mide 1,55 kg (de 67,48 a 69,03) frente a 1,06 kg con MAS. Los 36 intervalos de la tabla contienen el valor real.

Tabla 2.

Estimación puntual e intervalos de confianza en una muestra ilustrativa (n = 2 000) por diseño de muestreo

Método	Parámetro	Est. puntual	IC 90 %	IC 95 %	IC 99 %
MAS	Media del peso (kg)	68,56	[68,11; 69,01]	[68,03; 69,09]	[67,86; 69,26]
MAS	Desv. típica peso (kg)	12,20	[11,88; 12,52]	[11,82; 12,58]	[11,70; 12,69]
MAS	Proporción hombres	0,492	[0,474; 0,511]	[0,471; 0,514]	[0,464; 0,521]
Estratificado	Media del peso (kg)	68,65	[68,21; 69,08]	[68,13; 69,16]	[67,96; 69,33]
Estratificado	Desv. típica peso (kg)	12,45	[12,12; 12,77]	[12,06; 12,83]	[11,94; 12,95]
Estratificado	Proporción hombres	0,499	[0,481; 0,517]	[0,477; 0,521]	[0,470; 0,528]
Por racimos	Media del peso (kg)	68,25	[67,60; 68,90]	[67,48; 69,03]	[67,24; 69,27]
Por racimos	Desv. típica peso (kg)	12,27	[11,96; 12,59]	[11,89; 12,66]	[11,77; 12,77]
Por racimos	Proporción hombres	0,480	[0,459; 0,500]	[0,455; 0,504]	[0,447; 0,512]
Sistemático	Media del peso (kg)	68,49	[68,04; 68,93]	[67,96; 69,02]	[67,79; 69,18]
Sistemático	Desv. típica peso (kg)	12,12	[11,80; 12,43]	[11,74; 12,49]	[11,62; 12,61]
Sistemático	Proporción hombres	0,481	[0,463; 0,500]	[0,460; 0,503]	[0,453; 0,510]

Nota: resultados de una única muestra por diseño, con fines ilustrativos; no deben usarse para comparar la eficiencia de los diseños (véanse las Tablas 3 a 6). Valores reales: media del peso = 68,49 kg; desviación típica del peso = 12,14 kg; proporción de hombres = 0,490. IC: intervalo de confianza. Elaboración propia a partir de la simulación.

4.3 Sesgo y eficiencia en la estimación puntual

La Tabla 3 resume la estimación puntual sobre las 1 000 réplicas. Ningún diseño mostró sesgo apreciable. El mayor sesgo relativo fue de 0,197 %, en la proporción de hombres con MAS: el sesgo sin redondear es de +0,000965 sobre $P = 0,48998$, aunque la Tabla 3 lo muestra redondeado a +0,0010, lo que daría un 0,204 % si se calculara con las cifras redondeadas. En los 19 casos restantes el sesgo relativo no superó el 0,08 %, de modo que las diferencias de desempeño vienen de la varianza. La eficiencia, en cambio, varió mucho según el parámetro. En la media de la edad, que es la variable que define los estratos, el estratificado alcanzó una eficiencia relativa de 10,37 y redujo la RECM de 0,313 años con MAS a 0,097, con $Deff = 0,10$. La ganancia bajó a 2,94 en la desviación típica de la edad y casi desapareció en el peso (1,04 y 1,10) y en la proporción de hombres (1,00). La razón es que los estratos de edad explican el 89,7 % de la varianza de la edad, pero solo el 10,1 % de la del peso y prácticamente nada de la del sexo (3,4).

El sistemático siguió el mismo patrón con menor intensidad, gracias al orden del marco, con una eficiencia de 2,01 y un Deff de 0,50 en la media de la edad. Los conglomerados quedaron en el extremo opuesto, con Deff de 2,85 en la media del peso y de 1,94 en la de la edad, y eficiencias de 0,35 y 0,52 (13). En la Figura 2, las distribuciones del estratificado y del sistemático se concentran alrededor de la edad real, mientras que la de conglomerados es la más dispersa en ambas variables. La Figura 3 reúne la eficiencia relativa de todos los diseños y parámetros.

Tabla 3.

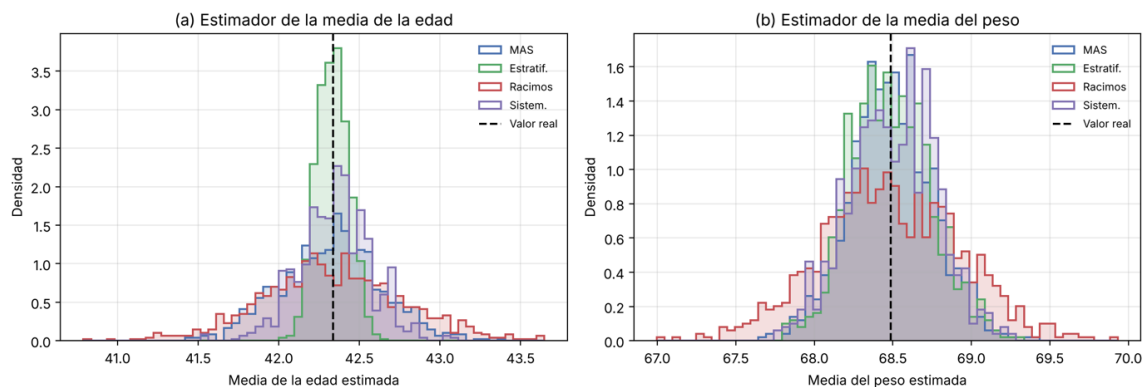
Desempeño de la estimación puntual sobre 1 000 réplicas: sesgo, sesgo relativo, raíz del error cuadrático medio (RECM), eficiencia relativa respecto al MAS y efecto de diseño

Método	Estimando	Sesgo	Sesgo rel. (%)	RECM	Ef. relativa	Deff
MAS	Media edad	-0,0079	-0,019	0,3132	1,00	1,00
MAS	DT edad	+0,0025	+0,018	0,2077	1,00	1,00
MAS	Media peso	+0,0061	+0,009	0,2611	1,00	1,00
MAS	DT peso	-0,0059	-0,048	0,1862	1,00	1,00
MAS	Prop. hombres	+0,0010	+0,197	0,0110	1,00	1,00
Estratificado	Media edad	-0,0021	-0,005	0,0972	10,37	0,10
Estratificado	DT edad	+0,0033	+0,024	0,1211	2,94	0,34
Estratificado	Media peso	-0,0001	0,000	0,2562	1,04	0,96
Estratificado	DT peso	+0,0030	+0,024	0,1776	1,10	0,91
Estratificado	Prop. hombres	-0,0001	-0,027	0,0110	1,00	1,01
Por racimos	Media edad	-0,0024	-0,006	0,4361	0,52	1,94
Por racimos	DT edad	-0,0110	-0,080	0,2012	1,07	0,94
Por racimos	Media peso	-0,0050	-0,007	0,4410	0,35	2,85
Por racimos	DT peso	-0,0021	-0,017	0,1873	0,99	1,01
Por racimos	Prop. hombres	-0,0003	-0,053	0,0125	0,77	1,31
Sistemático	Media edad	+0,0141	+0,033	0,2212	2,01	0,50
Sistemático	DT edad	+0,0051	+0,037	0,1666	1,55	0,64
Sistemático	Media peso	+0,0097	+0,014	0,2770	0,89	1,12
Sistemático	DT peso	-0,0011	-0,009	0,1722	1,17	0,86
Sistemático	Prop. hombres	-0,0002	-0,045	0,0114	0,93	1,08

Nota: el sesgo relativo es $100 \cdot \text{sesgo} / \theta$, calculado con valores sin redondear. La eficiencia relativa es $\text{ECMMAS} / \text{ECM}$; valores mayores que 1 indican mayor eficiencia que el muestreo aleatorio simple. El efecto de diseño (Deff) es el cociente de varianzas respecto al MAS; valores menores que 1 indican ganancia de precisión. Elaboración propia a partir de la simulación.

Figura 2.

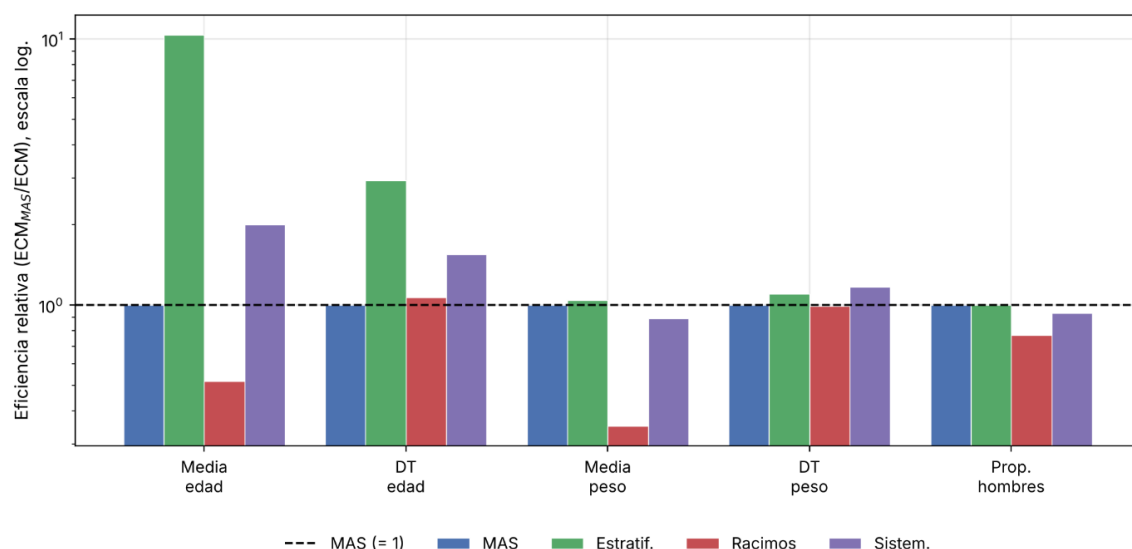
Distribuciones muestrales del estimador de la media de la edad (a) y del peso (b) bajo los cuatro diseños, sobre 1 000 réplicas. La línea discontinua marca el valor real



Nota: Elaboración propia a partir de la simulación (semilla 20260627).

Figura 3.

Eficiencia relativa ($ECMMAS/ECM$, escala logarítmica) por diseño y parámetro. La línea discontinua (= 1) corresponde al muestreo aleatorio simple



Nota: Elaboración propia a partir de la simulación (semilla 20260627).

4.4 Cobertura y amplitud de los intervalos de confianza

La Tabla 4 evalúa la validez y la precisión de los intervalos. Con MAS, las coberturas empíricas quedaron entre 89,0 % y 91,7 % al nivel del 90 %, entre 95,1 % y 96,6 % al 95 % y entre 98,8 % y 99,4 % al 99 %, lo que confirma que el procedimiento de construcción de intervalos funciona (Figura 4). Los conglomerados también cubrieron bien, entre 94,7 % y 97,1 % al 95 %, señal de que el error estándar calculado con la variabilidad entre racimos recoge la pérdida de precisión. El costo aparece en la amplitud, ya que el intervalo al 95 % de la media del peso promedió 1,72 kg, frente a 1,06 kg con MAS (Figura 5). Con el Deff observado de 2,85, igualar la precisión del MAS exigiría casi triplicar el tamaño de muestra (13).

El estratificado y el sistemático resultaron conservadores justo donde son más eficientes. La cobertura al 95 % de la desviación típica de la edad llegó a 99,9 % con el estratificado, y la de la media de la

edad, a 99,1 % con el sistemático. Estas coberturas son excesivas y no una virtud del diseño: el error estándar medio sobrestimó la dispersión real del estimador en un 80 % y un 39 %, respectivamente (cociente entre el error estándar medio y la desviación típica empírica del estimador de 1,80 y 1,39), de modo que los intervalos fueron más anchos de lo necesario y ocultan la ganancia de precisión. En el estratificado, la causa es que la fórmula de la ecuación (12) ignora la estratificación; en el sistemático, que la fórmula del MAS no recoge la estratificación implícita del marco ordenado (23,29). Por eso, en esos casos, la amplitud de los intervalos de la Tabla 4 no mide la precisión del diseño, que debe leerse en la RECM de la Tabla 3. La sección 4.6 muestra cómo cambia la cobertura con errores estándar coherentes con el diseño.

Tabla 4.

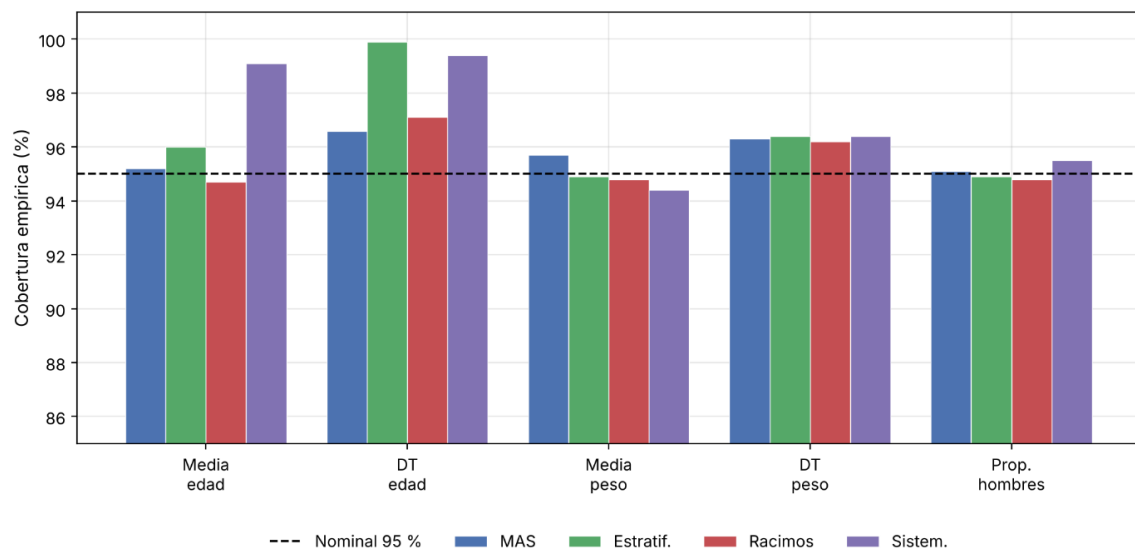
Cobertura empírica de los intervalos de confianza (90 %, 95 % y 99 %) y amplitud media del intervalo al 95 %, por diseño y estimando

Método	Estimando	Cob. 90 %	Cob. 95 %	Cob. 99 %	Amplitud IC 95 %
MAS	Media edad	89,0	95,2	98,8	1,2039
MAS	DT edad	91,7	96,6	99,4	0,8523
MAS	Media peso	90,4	95,7	99,1	1,0626
MAS	DT peso	91,3	96,3	99,4	0,7523
MAS	Prop. hombres	90,4	95,1	99,0	0,0438
Estratificado	Media edad	91,4	96,0	99,4	0,3871
Estratificado	DT edad	99,5	99,9	100,0	0,8524
Estratificado	Media peso	90,8	94,9	99,4	1,0075
Estratificado	DT peso	92,5	96,4	99,2	0,7529
Estratificado	Prop. hombres	90,3	94,9	98,8	0,0437
Por racimos	Media edad	89,3	94,7	98,1	1,7165
Por racimos	DT edad	92,4	97,1	99,2	0,8515
Por racimos	Media peso	89,9	94,8	99,0	1,7151
Por racimos	DT peso	90,4	96,2	99,1	0,7526
Por racimos	Prop. hombres	89,7	94,8	99,0	0,0486
Sistemático	Media edad	98,3	99,1	99,6	1,2041
Sistemático	DT edad	96,1	99,4	99,8	0,8525
Sistemático	Media peso	89,6	94,4	99,3	1,0630
Sistemático	DT peso	93,5	96,4	98,8	0,7526
Sistemático	Prop. hombres	88,6	95,5	99,1	0,0438

Nota: coberturas en porcentaje, calculadas con el error estándar principal (EE1); la amplitud se expresa en las unidades de cada parámetro (años, kg o proporción). Las coberturas de 99,9 % (estratificado, desviación típica de la edad) y 99,1 % (sistemático, media de la edad) reflejan errores estándar sobrestimados y no mayor validez; véase la Tabla 6. Elaboración propia a partir de la simulación.

Figura 4.

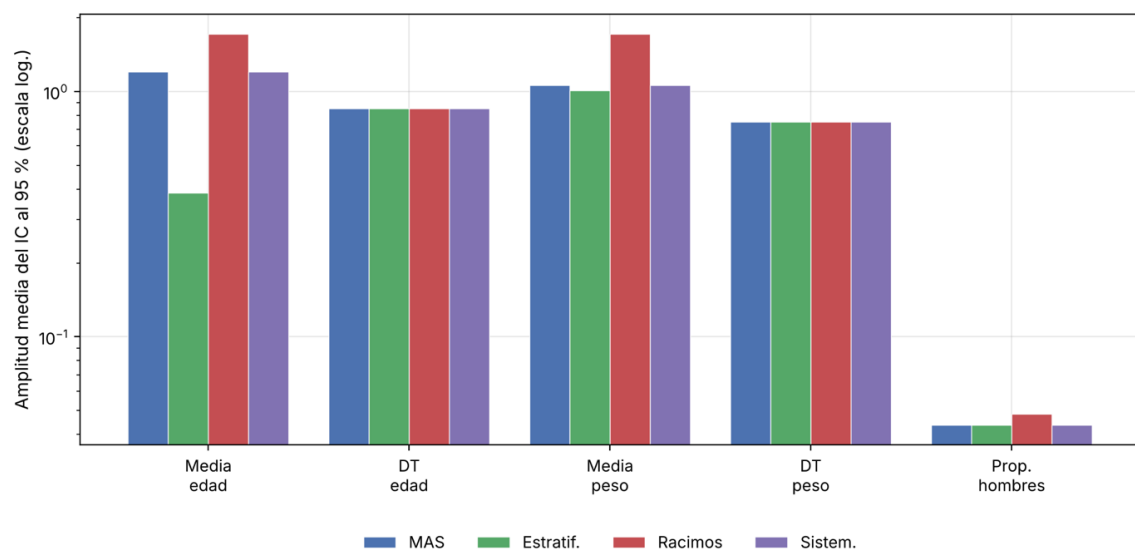
Cobertura empírica de los intervalos de confianza al 95 % por diseño y parámetro. La línea discontinua marca el nivel nominal del 95 %



Nota: Elaboración propia a partir de la simulación (semilla 20260627).

Figura 5.

Amplitud media de los intervalos de confianza al 95 % por diseño y parámetro (escala logarítmica). Con igual cobertura, un intervalo más estrecho indica mayor precisión



Nota: Elaboración propia a partir de la simulación (semilla 20260627).

4.5 Efecto de diseño y ordenamiento global de eficiencia

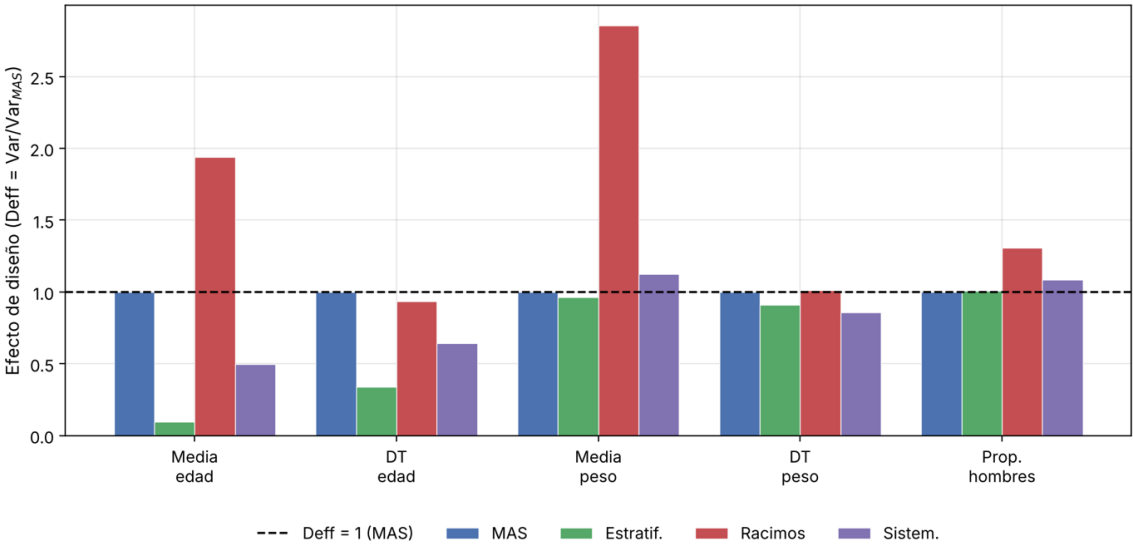
La Figura 6 muestra los efectos de diseño: claramente menores que 1 para el estratificado y el sistemático en los parámetros de la edad, cercanos a 1 en la mayoría de los demás casos y mayores que 1 para los conglomerados en las dos medias y en la proporción (1,94; 2,85 y 1,31). Al promediar la eficiencia relativa de los cinco estimandos (Tabla 5), el estratificado ocupa el primer lugar (3,29), seguido del sistemático (1,31), el MAS (1,00) y los conglomerados (0,74). Ese orden debe leerse con cautela, porque depende de los cinco estimandos elegidos y el promedio del estratificado está dominado por la eficiencia de 10,37 en la media de la edad. Sin ese estimando, su eficiencia media

<https://doi.org/10.71068/SIMJ0005>

baja a 1,52, y su mediana es 1,10, cerca de la del sistemático (1,14 y 1,17). El ordenamiento no expresa, por tanto, una superioridad general del estratificado, sino su ventaja en los parámetros asociados con la variable de estratificación, y podría cambiar con otros estimandos o con otra variable de estratificación. El promedio debe leerse además junto con la cobertura: con EE1, la ventaja del sistemático viene acompañada de la mayor desviación media de la cobertura, 2,20 puntos porcentuales, que se reduce a 0,50 con EE2 (sección 4.6).

Figura 6.

Efecto de diseño ($Deff = Var/Var_{MAS}$) por diseño y parámetro. Los valores por debajo de la línea (= 1) indican ganancia de precisión respecto al muestreo aleatorio simple



Nota: Elaboración propia a partir de la simulación (semilla 20260627).

Tabla 5.

Ordenamiento global de eficiencia de los cuatro diseños

Posición	Método	Ef. relativa media	Mediana de la Ef. relativa	Ef. relativa media sin media de la edad	Desv. media cob. 95 % con EE1 (pp)	Desv. media cob. 95 % con EE2 (pp)
1	Estratificado	3,29	1,10	1,52	1,50	0,56
2	Sistemático	1,31	1,17	1,14	2,20	0,50
3	MAS	1,00	1,00	1,00	0,78	0,58
4	Por racimos	0,74	0,77	0,79	0,80	0,24

Nota: la eficiencia relativa promedia o resume ECMMAS/ECM sobre los cinco estimandos; la columna «sin media de la edad» excluye ese estimando. La desviación media de la cobertura es el promedio de |cobertura empírica – 95| en puntos porcentuales (pp), con el error estándar principal (EE1) y con el alternativo (EE2). Elaboración propia a partir de la simulación. Todos los indicadores resumen eficiencia estadística a igual n; no incorporan costos operativos.

4.6 Sensibilidad de la cobertura al error estándar

La Tabla 6 compara la cobertura al 95 % obtenida con EE1 y con EE2 en los casos en que el error estándar principal no depende del diseño o es aproximado: las desviaciones típicas en los cuatro diseños y las medias y la proporción en el sistemático. Con la linealización, la cobertura de la desviación típica de la edad en el estratificado bajó de 99,9 % a 93,6 % y la amplitud media del intervalo, de 0,852 a 0,453 años; con diferencias sucesivas, la de la media de la edad en el sistemático pasó de 99,1 % a 95,4 % y la amplitud, de 1,204 a 0,857 años. Con EE2, el cociente entre el error estándar medio y la desviación típica empírica quedó entre 0,95 y 1,04 en todos los casos, y la desviación media de la cobertura respecto del 95 % bajó de 1,50 a 0,56 puntos en el estratificado y de 2,20 a 0,50 en el sistemático (Tabla 5). Con un error estándar coherente con el diseño, los intervalos de ambos diseños reflejan su ganancia de precisión.

La linealización también corrigió la leve sobrecobertura de la desviación típica con MAS (de 96,6 % a 93,9 % en la edad), que se explica porque la edad, truncada, tiene una curtosis menor que la normal (exceso de $-0,34$) y la ecuación (12) sobrestima entonces el error (34). Con $n = 2\,000$, sin embargo, la linealización quedó algo por debajo del nivel nominal en el MAS y el estratificado (entre 93,6 % y 94,8 %), por lo que no debe tomarse como una solución definitiva: los métodos de replicación, como el jackknife o el bootstrap, podrían calibrar mejor estos intervalos (29).

Tabla 6.

Análisis de sensibilidad: cobertura al 95 % y amplitud media del intervalo con el error estándar principal (EE1) y el alternativo (EE2)

Método	Estimando	EE/DT (EE1)	Cob. 95 % (EE1)	Amplitud (EE1)	EE/DT (EE2)	Cob. 95 % (EE2)	Amplitud (EE2)
MAS	DT edad	1,05	96,6	0,8523	0,95	93,9	0,7745
MAS	DT peso	1,03	96,3	0,7523	0,95	94,2	0,6967
Estratificado	DT edad	1,80	99,9	0,8524	0,95	93,6	0,4526
Estratificado	DT peso	1,08	96,4	0,7529	1,00	94,8	0,6944
Por racimos	DT edad	1,08	97,1	0,8515	1,01	95,1	0,7922
Por racimos	DT peso	1,02	96,2	0,7526	1,00	94,6	0,7335
Sistemático	DT edad	1,31	99,4	0,8525	1,04	95,5	0,6790
Sistemático	DT peso	1,11	96,4	0,7526	1,03	95,6	0,6961
Sistemático	Media edad	1,39	99,1	1,2041	0,99	95,4	0,8566
Sistemático	Media peso	0,98	94,4	1,0630	0,95	94,0	1,0333
Sistemático	Prop. hombres	0,98	95,5	0,0438	0,98	95,0	0,0438

Nota: EE/DT es el cociente entre el error estándar medio y la desviación típica empírica del estimador en las 1 000 réplicas (1 = error estándar sin sesgo). EE1: error estándar principal (ecuación 12 para la desviación típica; fórmula del MAS en el sistemático). EE2: linealización (ecuación 13) para la desviación típica y diferencias sucesivas (ecuación 14) en el sistemático. Cobertura en porcentaje; amplitud en las unidades del parámetro. Elaboración propia a partir de la simulación.

DISCUSIÓN

En una población conocida por completo, los resultados reproducen una idea central de la teoría del muestreo: ningún diseño es superior en todos los parámetros, y la eficiencia depende de cómo se relaciona el diseño con lo que se estima (3,4). El estratificado obtuvo la mayor eficiencia media, pero casi toda su ventaja estuvo en la edad, la variable que define los estratos; para el peso y la proporción de hombres se comportó como el MAS. Así lo anticipan la asignación de Neyman (1) y los textos de referencia (5,6,9), según los cuales la reducción de varianza depende de la fracción de la variabilidad que explican los estratos: aquí, el 89,7 % en la edad, el 10,1 % en el peso y prácticamente nada en el sexo.

La implicación práctica es que estratificar por una variable débilmente relacionada con el parámetro de interés no perjudica la precisión con asignación proporcional, pero tampoco la mejora, como muestran las eficiencias del peso (1,04) y de la proporción de hombres (1,00). En una encuesta con varios objetivos, la variable de estratificación debería elegirse en función del parámetro prioritario o combinar varias variables auxiliares; si los estratos se definen solo por una variable ajena a ese parámetro, el diseño añade complejidad operativa sin beneficio, y una asignación óptima para una variable puede ser ineficiente para las demás (24,25). Cuando la relación con la información auxiliar es débil o no lineal, los estimadores asistidos por modelos y la calibración, que usan esa información en la fase de estimación, ofrecen una vía complementaria (36,37).

Con los conglomerados pudimos contrastar la aproximación de Kish con lo observado. La correlación intraclase estimada en la población fue de 0,044 para la edad, 0,067 para el peso y 0,010 para el sexo; con $M = 25$, la ecuación (9) predice efectos de diseño de 2,04, 2,62 y 1,24, cercanos a los observados (1,94, 2,85 y 1,31). El resultado coincide con la experiencia en encuestas reales, donde el efecto de diseño por conglomeración rara vez es despreciable (13,28). La menor eficiencia estadística no implica necesariamente que el diseño deba descartarse: su atractivo práctico puede residir en la concentración territorial de las observaciones y en la disponibilidad de marcos de áreas, aspectos operativos que este experimento no cuantificó. Por consiguiente, ningún resultado autoriza concluir que los conglomerados sean más o menos eficientes económicamente. La investigación sobre muestreo espacialmente equilibrado muestra alternativas metodológicas para reducir la variabilidad bajo ciertas condiciones (38-40). La validez de los intervalos se mantuvo porque empleamos errores estándar basados en el diseño de conglomerados; con la fórmula del MAS se ignoraría la dependencia intrarracimo (5,14).

El sistemático mostró las dos caras de la estratificación implícita. Fue más eficiente que el MAS en los parámetros de la edad (2,01 en la media y 1,55 en la desviación típica), pero su varianza no puede estimarse sin sesgo con una sola muestra, y la fórmula del MAS que usamos en el análisis principal es

una aproximación conservadora (10,11,29). El estimador de diferencias sucesivas corrigió la sobrecobertura (Tabla 6), en línea con los trabajos que proponen estimadores aproximados para marcos ordenados (23). Conviene usar este diseño con marcos cuyo orden se conozca y, si es posible, con varios arranques aleatorios o con un estimador de diferencias sucesivas. En la proporción, el intervalo de Wald funcionó bien con $n = 2\,000$ y $P = 0,49$ (coberturas al 95 % entre 94,8 % y 95,5 %), aunque la literatura documenta su cobertura irregular en muestras pequeñas o con proporciones extremas, donde convienen los intervalos de Wilson o de Agresti y Coull (30-33).

Nos limitamos a diseños con probabilidades iguales. La teoría ofrece alternativas que aprovechan información auxiliar: el muestreo con probabilidades desiguales y el estimador de Horvitz y Thompson (2,9); los estimadores de razón, de regresión y otros que incorporan variables auxiliares (19,20); la calibración a totales conocidos (37); los estimadores asistidos por modelos de aprendizaje automático (36); los algoritmos de selección equilibrada y espacialmente dispersa (39,40), y el muestreo adaptativo por conglomerados para poblaciones raras o agregadas (41). El muestreo por conjuntos ordenados también ha mostrado ganancias sobre el MAS cuando se dispone de una variable de ordenamiento barata, incluso en diseños polietápicos (21,22). Nuestra comparación sirve como línea de base para esos diseños, y su patrón cualitativo coincide con lo que exponen los textos de referencia (5,6,14).

El estudio tiene limitaciones que restringen la generalización de las magnitudes obtenidas. Primero, las conclusiones proceden de una sola población sintética, construida bajo supuestos específicos (sección 3.1), y no de encuestas reales. Las cifras de eficiencia relativa y de efecto de diseño dependen de esa estructura, sobre todo de la correlación intraclase y de la asociación de cada variable con los criterios de estratificación u ordenamiento; otra población daría otras cifras, aunque el patrón cualitativo es el que anticipa la teoría (3). Segundo, usamos un solo tamaño de muestra ($n = 2\,000$), 1 000 réplicas por diseño y una única estructura de correlaciones entre edad, peso y sexo. Tercero, no modelamos errores ajenos al muestreo (no respuesta, medición, cobertura del marco), que en encuestas reales pueden dominar el error total (42); la no respuesta, por ejemplo, altera el desempeño relativo de los diseños (22). Además, la comparación mantuvo constante el número de unidades observadas y no el costo total de aplicar cada diseño, por lo que los resultados no identifican un diseño económicamente superior. El efecto favorable del estratificado está condicionado al papel de la edad en la construcción de estratos y no debe extrapolarse a otras variables o poblaciones.

Comparamos solo diseños de una etapa con probabilidades iguales, sin el muestreo polietápico, el de probabilidad proporcional al tamaño ni los estimadores asistidos por modelos, que son comunes en la práctica (4,9). Estratificamos por una sola variable con asignación proporcional; la asignación de Neyman o una estratificación multivariante podrían ampliar la ventaja observada (1,24). En el análisis principal, el intervalo de la desviación típica se basó en una aproximación normal de su error estándar que no depende del diseño, y el del sistemático, en la fórmula del MAS. El análisis de sensibilidad muestra que la linealización y las diferencias sucesivas corrigen la sobrecobertura, aunque la linealización queda algo por debajo del nivel nominal y otros métodos podrían calibrar mejor los intervalos (29). Por último, no sabemos todavía si el ordenamiento se mantiene con muestras pequeñas, con otros tamaños de racimo o con otras correlaciones, y ese análisis queda pendiente.

CONCLUSIONES

Sobre una población sintética de 1 000 000 de personas, conocida por completo y generada bajo supuestos específicos, los cuatro diseños dieron estimadores prácticamente insesgados, de modo que la elección entre ellos depende de la precisión del estimador y de la validez de los intervalos. Ningún diseño fue superior en todos los parámetros. El estratificado logró las mayores ganancias en la edad, la variable utilizada para formar los estratos; en el peso y en la proporción de hombres su ventaja prácticamente desapareció, y su eficiencia relativa media de 3,29 depende de los estimandos elegidos. El MAS mantuvo coberturas próximas a las nominales y el sistemático mejoró la precisión de la edad cuando se ordenó el marco con una variable auxiliar informativa, aunque la fórmula de varianza del MAS resultó conservadora. Los conglomerados fueron los menos eficientes estadísticamente a igual tamaño de muestra, con efectos de diseño de hasta 2,85. La cobertura de los intervalos dependió del estimador de error estándar, y las alternativas coherentes con el diseño atenuaron las sobrecoberturas observadas.

Estas conclusiones, obtenidas mediante una sola configuración de simulación, pueden orientar la planificación estadística, pero no sustituyen el análisis de encuestas reales. Conviene elegir el diseño en función del parámetro prioritario, de la información auxiliar disponible y de los supuestos de estratificación u ordenamiento. El estratificado es recomendable cuando existe una variable auxiliar informativa respecto del parámetro y se aplica una asignación adecuada; aquí su ventaja estuvo fuertemente ligada a la edad. El sistemático puede resultar útil con marcos de orden informativo, siempre que su incertidumbre se estime de forma apropiada. Los conglomerados requieren evaluar su efecto de diseño y podrían responder a necesidades logísticas, pero este trabajo no estima costos ni identifica superioridad económica. Para extender la comparación será necesario analizar otras poblaciones, tamaños de muestra, estructuras de correlación, diseños polietápicos, probabilidades desiguales, costos, no respuesta, errores de medición y métodos alternativos de construcción de intervalos.

REFERENCIAS BIBLIOGRÁFICAS

1. Neyman J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*. 1934;97(4):558-625. <https://doi.org/10.2307/2342192>.
2. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952;47(260):663-685. <https://doi.org/10.1080/01621459.1952.10483446>.
3. Cochran WG. *Sampling techniques*. 3rd ed. New York: John Wiley & Sons; 1977.
4. Särndal CE, Swensson B, Wretman J. *Model assisted survey sampling*. New York: Springer-Verlag; 1992. <https://doi.org/10.1007/978-1-4612-4378-6>
5. Lohr SL. *Sampling: design and analysis*. 3rd ed. Boca Raton: Chapman and Hall/CRC; 2022. <https://doi.org/10.1201/9780429298899>
6. Kim JK. *Statistics in survey sampling*. Boca Raton: Chapman and Hall/CRC; 2025. <https://doi.org/10.1201/9781003606147>

7. Kalton G. Probability vs. nonprobability sampling: from the birth of survey sampling to the present day. *Statistics in Transition New Series*. 2023;24(3):1-22.
<https://doi.org/10.59170/stattrans-2023-029>.
8. Wu C. Statistical inference with non-probability survey samples. *Survey Methodology*. 2022;48(2):283-311. <https://www150.statcan.gc.ca/n1/pub/12-001-x/2022002/article/00002-eng.htm>.
9. Tillé Y. Sampling and estimation from finite populations. Hoboken: John Wiley & Sons; 2020.
<https://doi.org/10.1002/9781119071259>
10. Madow WG, Madow LH. On the theory of systematic sampling, I. *The Annals of Mathematical Statistics*. 1944;15(1):1-24. <https://doi.org/10.1214/aoms/1177731312>.
11. Iachan R. Systematic sampling: a critical review. *International Statistical Review*. 1982;50(3):293-303. <https://doi.org/10.2307/1402499>.
12. Azeem M, Hussain S, Ijaz M, Salahuddin N, Salam A. An improved version of systematic sampling design for use with linear trend data. *Heliyon*. 2023;9(6):e17121.
<https://doi.org/10.1016/j.heliyon.2023.e17121>.
13. Kish L. Survey sampling. New York: John Wiley & Sons; 1965.
14. Valliant R, Dever JA, Kreuter F. Practical tools for designing and weighting survey samples. 2nd ed. Cham: Springer; 2018. <https://doi.org/10.1007/978-3-319-93632-1>
15. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 2019;38(11):2074-2102. <https://doi.org/10.1002/sim.8086>.
16. Chapuis K, Taillandier P, Drogoul A. Generation of synthetic populations in social simulations: a review of methods and practices. *Journal of Artificial Societies and Social Simulation*. 2022;25(2):6. <https://doi.org/10.18564/jasss.4762>.
17. Siepe BS, Bartoš F, Morris TP, Boulesteix AL, Heck DW, Pawel S. Simulation studies for methodological research in psychology: a standardized template for planning, preregistration, and reporting. *Psychological Methods*. 2024. Publicación anticipada en línea.
<https://doi.org/10.1037/met0000695>
18. Kondratev M, Corrao MV, Armstrong R, Smith AMS. Assessing the uncertainty of traditional sample-based forest inventories in mixed and single species conifer systems using a digital forest twin. *Forests*. 2025;16(11):1617. <https://doi.org/10.3390/f16111617>.
19. Daraz U, Wu J, Agustiana D, Emam W. Finite population variance estimation using Monte Carlo simulation and real life application. *Symmetry*. 2025;17(1):84.
<https://doi.org/10.3390/sym17010084>.
20. Alghamdi AS, Almulhim FA. Optimizing finite population mean estimation using simulation and empirical data. *Mathematics*. 2025;13(10):1635. <https://doi.org/10.3390/math13101635>.
21. Abbas M, Shehzad MA, Rabia M, Khurram H, Ijaz M. Estimation of finite population mean in a complex survey sampling. *PLoS ONE*. 2025;20(5):e0324559.
<https://doi.org/10.1371/journal.pone.0324559>.
22. Rehman SA, Al-Essa LA, Shabbir J, Khan Z. On the efficiency of paired ranked set sampling for estimating the population mean in the presence of non-response. *Scientific Reports*. 2024;14:25334. <https://doi.org/10.1038/s41598-024-76072-y>.

23. Franceschi S, Fattorini L, Gregoire TG. Exploiting nearest-neighbour maps for estimating the variance of sample mean in equal-probability systematic sampling of spatial populations. *Spatial Statistics*. 2024;64:100865. <https://doi.org/10.1016/j.spasta.2024.100865>.
24. Wesołowski J, Wieczorkowski R, Wójciak W. Recursive Neyman algorithm for optimum sample allocation under box constraints on sample sizes in strata. *Survey Methodology*. 2024;50(2):487-511. <https://www150.statcan.gc.ca/n1/pub/12-001-x/2024002/article/00003-eng.htm>.
25. Sauer S, Hedt-Gauthier B, Haneuse S. Practical strategies for operationalizing optimal allocation in stratified cluster-based outcome-dependent sampling designs. *Statistics in Medicine*. 2023;42(7):917-935. <https://doi.org/10.1002/sim.9650>.
26. Lohmann A, Astivia OLO, Morris TP, Groenwold RHH. It's time! Ten reasons to start replicating simulation studies. *Frontiers in Epidemiology*. 2022;2:973470. <https://doi.org/10.3389/fepid.2022.973470>.
27. Pawel S, Kook L, Reeve K. Pitfalls and potentials in simulation studies: questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Biometrical Journal*. 2024;66(1):2200091. <https://doi.org/10.1002/bimj.202200091>.
28. Rhoda DA, Soofi SB, Hussain I, Clary CB, Yu M, Khanyile S, et al. Sample size parameters for vaccination coverage cluster surveys: insights from Pakistan. *Journal of Survey Statistics and Methodology*. 2025;13(5):587-614. <https://doi.org/10.1093/jssam/smaf033>.
29. Wolter KM. Introduction to variance estimation. 2nd ed. New York: Springer; 2007. <https://doi.org/10.1007/978-0-387-35099-8>
30. Wilson EB. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*. 1927;22(158):209-212. <https://doi.org/10.1080/01621459.1927.10502953>.
31. Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*. 1998;52(2):119-126. <https://doi.org/10.1080/00031305.1998.10480550>.
32. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statistical Science*. 2001;16(2):101-133. <https://doi.org/10.1214/ss/1009213286>.
33. Andersson PG. The Wald confidence interval for a binomial p as an illuminating “bad” example. *The American Statistician*. 2023;77(4):443-448. <https://doi.org/10.1080/00031305.2023.2183257>.
34. Curto J. Confidence intervals for means and variances of nonnormal distributions. *Communications in Statistics - Simulation and Computation*. 2023;52(9):4414-4430. <https://doi.org/10.1080/03610918.2021.1963448>.
35. Bhushan S, Pandey S. Optimality of novel estimators for variance estimation under successive sampling. *Communications in Statistics - Theory and Methods*. 2025;54(14):4635-4653. <https://doi.org/10.1080/03610926.2024.2425746>.
36. Dagdou M, Goga C, Haziza D. Model-assisted estimation through random forests in finite population sampling. *Journal of the American Statistical Association*. 2023;118(542):1234-1251. <https://doi.org/10.1080/01621459.2021.1987250>.
37. Deville JC, Särndal CE. Calibration estimators in survey sampling. *Journal of the American Statistical Association*. 1992;87(418):376-382. <https://doi.org/10.1080/01621459.1992.10475217>.

38. Heikkinen J, Henttonen HM, Katila M, Tuominen S. Stratified, spatially balanced cluster sampling for cost-efficient environmental surveys. *Environmetrics*. 2025;36(5):e70019. <https://doi.org/10.1002/env.70019>.
39. Prentius W, Grafström A. How to find the best sampling design: a new measure of spatial balance. *Environmetrics*. 2024;35(7):e2878. <https://doi.org/10.1002/env.2878>.
40. Jauslin R, Panahbehagh B, Tillé Y. Sequential spatially balanced sampling. *Environmetrics*. 2022;33(8):e2776. <https://doi.org/10.1002/env.2776>.
41. Yasmeen U, Noor-ul-Amin M, Hanif M. Variance estimation in stratified adaptive cluster sampling. *Statistics in Transition New Series*. 2022;23(1):173-184. <https://doi.org/10.2478/stattrans-2022-0010>.
42. Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. *Survey methodology*. 2nd ed. Hoboken: John Wiley & Sons; 2009.

DISPONIBILIDAD DE DATOS Y CÓDIGO

No se utilizaron datos primarios procedentes de participantes. Se adjuntan dos scripts Python de comprobación independiente: S1 (Suplementario_S1_simulacion_muestreo.py) implementa el esquema de generación poblacional, selección de muestras y medidas de desempeño descrito en el método; S2 (Suplementario_S2_figuras.py) genera gráficos auxiliares a partir de sus resultados. Ambos archivos son ejecutables y documentan sus parámetros. No sustituyen los scripts originales con los que se produjeron las tablas y figuras del manuscrito, que no estuvieron disponibles para esta revisión; por ello, la reproducción numérica exacta de los resultados publicados continúa pendiente de contrastación por el autor.

CONSIDERACIONES ÉTICAS

El estudio se desarrolló mediante una revisión de literatura basada exclusivamente en fuentes documentales previamente publicadas. No involucró la participación directa de seres humanos o animales, la realización de intervenciones, el uso de muestras biológicas ni el acceso a datos personales identificables. En consecuencia, no requirió aprobación de un comité de ética ni la obtención de consentimiento informado. Durante el proceso investigativo se observaron los principios de integridad científica, atribución adecuada de la autoría, citación rigurosa y uso responsable de las fuentes consultadas.

DECLARACIÓN DE USO DE LA INTELIGENCIA ARTIFICIAL (IA)

Los autores afirman que la herramienta de inteligencia artificial generativa se utilizó exclusivamente para apoyar la organización del artículo y la traducción del texto. Su uso no determinó el contenido científico, el diseño del estudio, la selección de fuentes, el análisis de datos, la interpretación de resultados ni las conclusiones. La responsabilidad total del contenido sigue siendo de los autores.

FINANCIACIÓN

Los autores declaran que este estudio no recibió ningún tipo de financiación externa por parte de agencias públicas, privadas, ni de organizaciones sin ánimo de lucro. Todas las actividades de investigación, análisis y desarrollo fueron realizadas con recursos propios.

CONFLICTO DE INTERESES

<https://doi.org/10.71068/SIMJ0005>

Los autores declaran que no tienen conflictos de intereses relacionados con este estudio y que todos los procedimientos seguidos cumplen con los estándares éticos establecidos por la revista. Asimismo, confirman que este trabajo es inédito y no ha sido publicado, ni parcial ni totalmente, en ninguna otra publicación.

CONTRIBUCIÓN DEL AUTOR

Nombres e iniciales de los autores: Marco Antonio Ramos Espinal (MARE)

- 1. Conceptualización:** MARE
- 2. Curación de datos:** MARE
- 3. Análisis formal:** MARE
- 4. Adquisición de financiación:** no aplicable
- 5. Investigación:** MARE
- 6. Metodología:** MARE
- 7. Administración de proyectos:** MARE
- 8. Recursos:** no aplicable
- 9. Software:** no aplicable
- 10. Supervisión:** MARE
- 11. Validación:** MARE
- 12. Vista:** MARE
- 13. Borrador original:** MARE
- 14. Redacción, revisión y edición:** MARE