



TRABAJO FIN DE MÁSTER EN BIOESTADÍSTICA

# IMPUTACIÓN MÚLTIPLE Y VALIDACIÓN BOOTSTRAP EN MODELOS PRONÓSTICOS

SEPTIEMBRE 2022

Alumna: Melina Peressini Álvarez

Director: Borja Manuel Fernández Félix

Tutora académica: María Jesús López Herrero



## **Agradecimientos**

A Borja y a María Jesús, por escucharme con paciencia y ayudarme a lo largo de este trabajo.

A todas las personas de la Unidad de Bioestadística del Hospital Ramón y Cajal, por acogerme de forma cálida y dejarme aprender de ellas.

A mis profesores, por ser parte de este camino que recorro ahora.

A las personas que me acompañan, a las que me han acompañado; por tantas cosas que atesoro con profundo cariño.



## Índice

1. Resumen.....	1
2. Abstract .....	2
3. Introducción .....	3
4. Objetivos .....	7
5. Metodología .....	8
5.1. Validación bootstrap .....	8
5.2. Valores perdidos .....	10
5.2.1. Mecanismos de pérdida .....	11
5.2.2. Imputación múltiple .....	12
5.2.2.1. Métodos univariantes de imputación múltiple.....	14
5.2.2.1.1. <i>Predictive mean matching</i> .....	14
5.2.2.1.2. Regresión logística bayesiana .....	15
5.2.2.2. Imputación multivariante por ecuaciones encadenadas (MICE) .....	16
5.2.2.3. Valores perdidos en la variable de interés en el modelo pronóstico: enfoques MI y MID	17
5.3. Integración de la imputación múltiple en la validación bootstrap.....	18
5.3.1. Estrategia MI-BS .....	19
5.3.2. Estrategia BS-MI .....	20
5.4. Regresión logística binomial .....	21
5.5. Medidas de rendimiento predictivo.....	22
5.5.1. Discriminación .....	22
5.5.1.1. Estadístico C.....	23
5.5.2. Calibración.....	23
5.5.2.1. <i>Calibration-in-the-large</i> (CITL) .....	24
5.5.2.2. <i>Calibration slope</i> .....	25
6. Planteamiento del problema: estudio de simulación.....	27
6.1. Descripción de los escenarios del estudio .....	29
6.2. Generación de los datos .....	29
6.2.1. Obtención de muestras completas .....	30
6.2.2. Introducción de valores perdidos.....	31
6.2.2.1. Mecanismo MCAR .....	31
6.2.2.2. Mecanismo MAR.....	31
6.3. Imputación múltiple.....	33
6.4. Modelización y medidas de rendimiento predictivo.....	33

6.5.	Estrategias a comparar .....	33
6.6.	Evaluación de las estrategias .....	34
6.7.	Software.....	35
7.	Resultados.....	37
7.1.	Rendimiento predictivo.....	37
7.1.1.	Rendimiento de referencia o <i>gold standard</i> .....	37
7.1.2.	Rendimiento aparente .....	38
7.1.3.	Estimación del rendimiento a través de las estrategias de validación .....	38
7.1.3.1.	Discriminación .....	39
7.1.3.2.	Calibración.....	42
7.1.3.2.1.	<i>Calibration slope</i> .....	42
7.1.3.2.2.	<i>Calibration-in-the-large</i> (CITL).....	45
7.2.	Tiempo computacional .....	48
8.	Discusión .....	52
9.	Conclusiones .....	54
10.	Bibliografía.....	55
11.	Anexos.....	58
	Anexo 1. Estadístico C de referencia.....	58
	Anexo 2. <i>Calibration slope</i> de referencia .....	60
	Anexo 3. <i>Calibration-in-the-large</i> de referencia .....	62
	Anexo 4. Diferencia $\Delta$ estadístico C sin validación interna .....	64
	Anexo 5. Diferencia $\Delta$ <i>calibration slope</i> sin validación interna .....	66
	Anexo 6. Diferencia $\Delta$ CITL sin validación interna .....	68
	Anexo 7. Diferencia $\Delta$ estadístico C por estrategias: gráficos de caja bajo MCAR.....	70
	Anexo 8. Diferencia $\Delta$ estadístico C por estrategias: valores medios bajo MCAR.....	71
	Anexo 9. Diferencia $\Delta$ <i>calibration slope</i> por estrategias: gráficos de caja bajo MCAR.....	72
	Anexo 10. Diferencia $\Delta$ <i>calibration slope</i> por estrategias: valores medios bajo MCAR.....	73
	Anexo 11. Diferencia $\Delta$ CITL por estrategias: gráficos de caja bajo MCAR.....	74
	Anexo 12. Diferencia $\Delta$ CITL por estrategias: valores medios bajo MCAR .....	75
	Anexo 13. Diferencia $\Delta$ estadístico C por estrategias: estadísticos descriptivos bajo MCAR	76
	Anexo 14. Diferencia $\Delta$ estadístico C por estrategias: estadísticos descriptivos bajo MAR ..	81
	Anexo 15. Diferencia $\Delta$ <i>calibration slope</i> por estrategias: estadísticos descriptivos bajo MCAR .....	86
	Anexo 16. Diferencia $\Delta$ <i>calibration slope</i> por estrategias: estadísticos descriptivos bajo MAR .....	91

Anexo 17. Diferencia $\Delta$ CITL por estrategias: estadísticos descriptivos bajo MCAR .....	96
Anexo 18. Diferencia $\Delta$ CITL por estrategias: estadísticos descriptivos bajo MAR.....	101
Anexo 19. Código R para la simulación .....	106

## 1. Resumen

En el ámbito biomédico, los modelos pronósticos se emplean habitualmente para predecir la probabilidad de que un paciente presente una determinada condición. Su validación interna es necesaria para estimar su rendimiento predictivo en nuevos individuos, y puede llevarse a cabo empleando la técnica de remuestreo bootstrap. Ante la presencia de valores perdidos, las técnicas estadísticas clásicas requieren su tratamiento previo, que puede abordarse mediante imputación múltiple: (I) los valores perdidos se imputan múltiples veces, (II) el análisis estadístico se realiza en cada una de las muestras completas resultantes y (III) las estimaciones obtenidas para el parámetro de interés se combinan. En el marco de la validación interna bootstrap, la forma en que la imputación múltiple debe integrarse en el proceso de remuestreo se encuentra actualmente en estudio. En el presente trabajo, se realiza un estudio de simulación para evaluar diferentes estrategias cuando se tienen valores perdidos tanto en los predictores como en la variable de interés de un modelo logístico. En la estrategia MI-BS, se aplica en primer lugar la imputación múltiple y el remuestreo se realiza sobre cada una de las muestras imputadas. En la estrategia BS-MI, se realiza en primer lugar el remuestreo y la imputación múltiple se aplica sobre cada una de las muestras bootstrap. La estrategia BS-MI proporciona estimadores de rendimiento de menor sesgo en la práctica mayoría de los escenarios estudiados. Las diferencias entre estrategias se encuentran cuando el número de eventos por variable (EPV) es reducido y se desdibujan conforme éste aumenta.

Palabras clave: modelos pronósticos, validación interna, bootstrap, imputación múltiple, MICE

## 2. Abstract

In the biomedical field, prognostic models are commonly used to predict the probability of a patient having a certain condition. Internal validation is necessary to estimate their predictive performance in new individuals, and can be carried out using the bootstrap resampling technique. Missing data need to be handled prior to classical statistical analyses. Multiple imputation is a popular approach to addressing the presence of missing data: (I) missing values are imputed several times, (II) statistical analysis is performed on each of the resulting complete samples, and (III) the estimates obtained for the parameter of interest are pooled. In the framework of bootstrap internal validation, the way multiple imputation of missing values and resampling should be combined is currently under study. In this work, a simulation study is performed to evaluate different strategies when missing values are in both covariates and outcome of a logistic model. In the MI-BS strategy, multiple imputation is applied first and bootstrap resampling is performed on each of the complete samples. On the other hand, in the BS-MI strategy resampling is performed first and multiple imputation is applied on each of the bootstrap samples. BS-MI provides less biased estimators in most of the scenarios. Differences between both strategies are found when the number of events per variable (EPV) is low, and become smaller as it increases.

Keywords: prognostic models, internal validation, bootstrap, multiple imputation, MICE

### 3. Introducción

En el ámbito biomédico, la probabilidad de ocurrencia de un evento o condición asociada al paciente se encuentra en la base de numerosas decisiones llevadas a cabo por los profesionales. Puede considerarse para diagnosticar una patología, para descartarla, o para determinar la necesidad de realizar pruebas adicionales de cara a establecer un diagnóstico. En relación con el pronóstico de una enfermedad presentada por el paciente, las probabilidades asociadas pueden guiar tanto la toma de decisiones terapéuticas, como la información trasladada al paciente y a su familia.

Para predecir la probabilidad de ocurrencia del evento de interés suelen considerarse diversas características del paciente, la sintomatología que presenta, resultados de distintas pruebas médicas que se le han realizado, etc.; siendo habitual el empleo de modelos de regresión logística o de regresión Cox de carácter multivariable.

En este ámbito, los modelos predictivos reciben la denominación de *modelos pronósticos* (1). Los modelos pronósticos se desarrollan a partir de una muestra de individuos procedentes de una población determinada, con el objetivo de proporcionar predicciones válidas para nuevos individuos de dicha población. Los modelos pronósticos no siempre funcionan bien en la práctica y, por lo tanto, deben ser previamente validados. La validación de un modelo pronóstico supone estimar su rendimiento predictivo en nuevos individuos, constituyendo un aspecto fundamental del proceso de modelización predictiva. En la validación de los modelos pronósticos, se distingue entre *validación interna* y *validación externa*.

A través de la validación interna de un modelo pronóstico, se estima su rendimiento predictivo en individuos que proceden de la misma población que aquellos que conforman la muestra de desarrollo. Se lleva a cabo empleando esta misma muestra, mediante técnicas tales como *split sample*, *validación cruzada* y *validación bootstrap*. La *validez interna* de un modelo pronóstico hace referencia a su *reproductibilidad*.

En la validación externa de un modelo pronóstico, se estima su rendimiento predictivo en individuos pertenecientes a poblaciones relacionadas con aquella para la que el modelo fue

inicialmente desarrollado, pero distintas a ésta en algún aspecto (2). Para ello, es necesario contar con muestras de individuos procedentes de estas otras poblaciones. Casos particulares de la validación externa son la *validación temporal*, que se refiere al rendimiento predictivo que presenta un modelo pronóstico desarrollado con anterioridad en una cohorte de individuos más reciente o actual; y la *validación geográfica*, que se refiere al rendimiento predictivo que presenta un modelo pronóstico cuando se emplea en individuos procedentes de otras regiones. La *validez externa* de un modelo pronóstico tiene que ver con la generalización o extensión de su uso.

La declaración TRIPOD (*Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis*) establece que los estudios en que se desarrollan nuevos modelos pronósticos deben incluir siempre su validación interna (1).

La validación bootstrap es una técnica de validación interna en la que se obtiene el rendimiento aparente del modelo pronóstico (es decir, el que presenta sobre la misma muestra de individuos con que ha sido desarrollado) y, a continuación, se ofrece una estimación del *optimismo* presente en este rendimiento aparente. Para la estimación del optimismo, se hace uso del método de *remuestreo bootstrap*, mediante el que se generan múltiples *muestras bootstrap* a partir de la muestra de desarrollo. Finalmente, el rendimiento del modelo pronóstico sobre nuevos individuos se estima corrigiendo el rendimiento aparente por el optimismo estimado.

Los estimadores bootstrap de rendimiento han mostrado empíricamente un sesgo inferior al de estimadores correspondientes a otras técnicas de validación interna y una variabilidad próxima a la del rendimiento aparente (3,4).

Los valores perdidos constituyen un problema frecuente en estudios epidemiológicos y clínicos, requiriendo ser tratados de forma previa a la aplicación de técnicas estadísticas clásicas. Habitualmente, el tratamiento de valores perdidos se aborda mediante técnicas de imputación múltiple, que involucran tres etapas (5):

- (I) Los valores perdidos se imputan (es decir, se reemplazan por valores plausibles)  $M$  veces ( $M > 1$ ), obteniéndose  $M$  versiones completas de la muestra de desarrollo.
- (II) El análisis estadístico se lleva a cabo en cada una de las  $M$  muestras completas, obteniéndose  $M$  estimaciones para cada uno de los parámetros de interés.
- (III) Estas  $M$  estimaciones se combinan, recogiendo la incertidumbre correspondiente a la imputación de los valores perdidos.

La presencia de valores perdidos en la muestra de desarrollo añade complejidad al proceso de validación interna de un modelo pronóstico. En el marco de la validación bootstrap, la forma en que la imputación múltiple debe integrarse en el proceso de remuestreo para la estimación del optimismo no resulta clara.

Diversas estrategias son posibles: puede llevarse a cabo, en primer lugar, la imputación múltiple de valores perdidos sobre la muestra de desarrollo y, a continuación, realizarse el remuestreo bootstrap sobre cada una de las  $M$  muestras imputadas; alternativamente, puede realizarse, en primer lugar, el remuestreo bootstrap sobre la muestra de desarrollo y, posteriormente, llevarse a cabo la imputación múltiple de valores perdidos sobre cada una de las muestras bootstrap generadas.

Estudiar tales estrategias de combinación resulta de fundamental importancia, siendo deseable que, en presencia de valores perdidos, los estimadores bootstrap de rendimiento corregido conserven sus buenas propiedades.

Musoro et al. (6) estudiaron la integración de la imputación múltiple en la validación bootstrap en el marco de la modelización de una variable continua mediante regresión LASSO (*Least Absolute Shrinkage and Selection Operator*). Wahl et al. (7) estudiaron la integración de la imputación múltiple en diversas técnicas de validación interna, incluyendo la validación bootstrap, en modelos de regresión logística binomial.

Estos trabajos concluyen que realizar el remuestreo sobre la muestra de desarrollo con valores perdidos, llevando a cabo la imputación múltiple sobre las muestras bootstrap, da lugar a estimadores de rendimiento de menor sesgo; por el contrario, abordar inicialmente la imputación múltiple de valores perdidos, llevando a cabo el remuestreo bootstrap sobre las muestras imputadas, conduce a una mayor sobrestimación del rendimiento del modelo pronóstico.

En ambas investigaciones, se considera la presencia de valores perdidos únicamente en los predictores del modelo pronóstico, quedando sin abordar el estudio de la problemática cuando la variable de interés también se presenta incompleta.

#### 4. Objetivos

El presente trabajo tiene como objetivo principal la realización de un estudio de simulación que permita comparar diferentes estrategias de integración de la imputación múltiple en la validación bootstrap de modelos de regresión logística binomial, en presencia de valores perdidos tanto en los predictores como en la variable correspondiente al evento de interés.

Como objetivo secundario, se plantea evaluar la complejidad temporal que presenta la aplicación de cada una de estas estrategias en el marco del estudio de simulación realizado.

## 5. Metodología

En la presente sección se describen los procedimientos y métodos empleados para la realización del trabajo.

### 5.1. Validación bootstrap

El rendimiento predictivo que un modelo pronóstico presenta sobre la misma muestra de individuos con que se desarrolla se denomina *rendimiento aparente*. Generalmente, este rendimiento aparente es superior al que el modelo presenta cuando se evalúa sobre nuevos individuos, incluso si estos proceden de la misma población.

Esta diferencia, a la que se denomina *optimismo*, suele presentarse de forma más acusada cuanto mayor es la complejidad del modelo pronóstico y cuanto menor es el tamaño de la muestra empleada para desarrollarlo. En la modelización de variables de naturaleza dicotómica, es clave el concepto de *eventos por variable* (EPV): considerando como número de eventos el mínimo entre el número de individuos que experimentan el evento y el de individuos que no lo hacen, el EPV se obtiene dividiendo este valor entre los grados de libertad requeridos para representar todas las covariables del modelo pronóstico (8). Valores reducidos de EPV se relacionan con la introducción de sesgo en la estimación de los coeficientes del modelo desarrollado y con un incremento del *sobreajuste* (*overfitting*) (9,10).

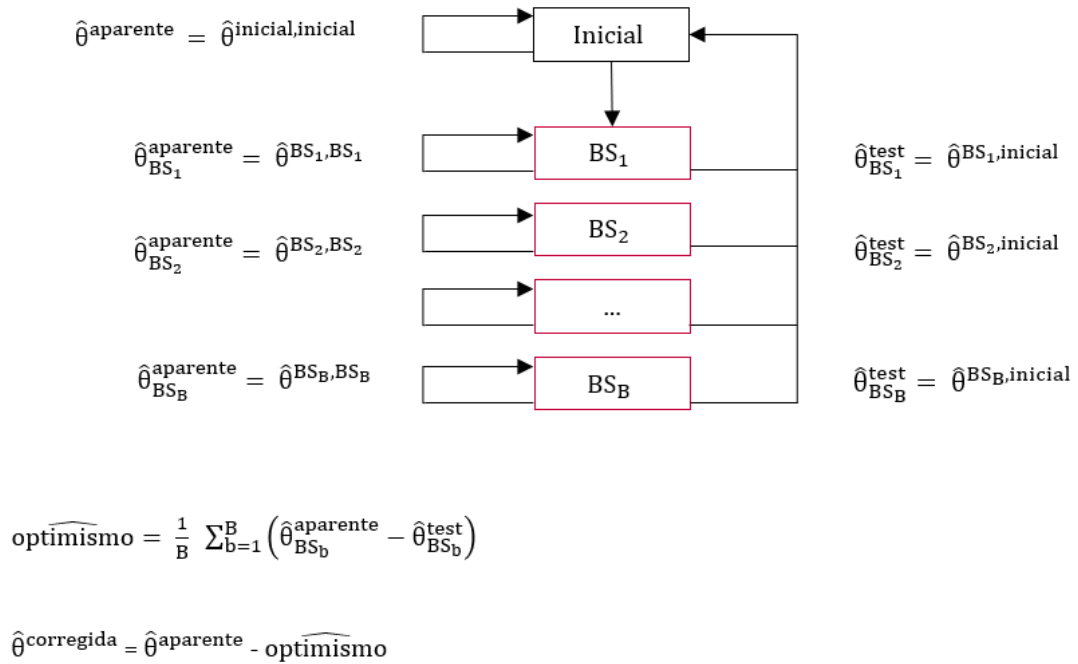
La validación bootstrap es una técnica de validación interna en la que el rendimiento aparente del modelo desarrollado se corrige a través de una estimación del optimismo. A diferencia de otras técnicas de validación interna, en las que parte de la muestra se destina exclusivamente a la validación (*split sample*), la validación bootstrap permite aprovechar todas las observaciones de la muestra para el desarrollo del modelo pronóstico. Para la estimación del optimismo, se emplea el método de remuestreo bootstrap, a través del cual se replica el proceso de toma de muestras en una población (11):

- (I) A partir de la muestra inicial, se generan **B** muestras bootstrap del mismo tamaño que esta mediante muestreo aleatorio con reemplazamiento. El empleo de

reemplazamiento en el remuestreo hace que, en cada muestra bootstrap, determinados individuos puedan aparecer más de una vez, mientras que otros pueden no aparecer ninguna.

- (II) En cada muestra bootstrap, se reproduce el proceso llevado a cabo para el desarrollo del modelo pronóstico, utilizando las mismas técnicas y estrategias; esto incluye los métodos de selección de predictores, si se han empleado (4).
- (III) El modelo desarrollado en cada muestra bootstrap se evalúa en esa misma muestra y en la muestra inicial. El rendimiento obtenido sobre la misma muestra bootstrap, denominado *rendimiento aparente bootstrap*, representa una estimación del rendimiento aparente; mientras que el rendimiento obtenido sobre la muestra inicial, denominado *rendimiento test bootstrap*, representa una estimación del rendimiento sobre nuevos individuos.
- (IV) Para estimar el optimismo, en cada muestra bootstrap se obtiene la diferencia entre el rendimiento aparente bootstrap y el rendimiento test bootstrap. Finalmente, se toma el valor promedio de las **B** diferencias correspondientes a las **B** muestras bootstrap generadas. Para obtener estimaciones estables del optimismo, se recomienda generar al menos **B = 100** muestras bootstrap (11).
- (V) La estimación del rendimiento predictivo del modelo pronóstico en nuevos individuos se obtiene sustrayendo la estimación del optimismo a su rendimiento aparente (2,4).

Los estimadores de rendimiento obtenidos de esta manera han mostrado empíricamente ser adecuados, presentando un sesgo próximo a cero y una variabilidad próxima a la del rendimiento aparente (3,4). El procedimiento se representa en la Figura 1.



**Figura 1.** Validación bootstrap. Se sigue la notación empleada por Wahl et al. (7) :  $\theta$  denota a la medida de rendimiento de interés; el primer superíndice denota la muestra empleada para el desarrollo del modelo y el segundo superíndice denota la muestra sobre la que se evalúa el modelo.

Lógicamente, las observaciones que integran las muestras bootstrap no son independientes de aquellas que conforman la muestra inicial. Asintóticamente, una muestra bootstrap incluye, en media, a un 63.2% de los individuos de la muestra inicial (11).

En la literatura especializada pueden encontrarse variantes del procedimiento, conocidas como *bootstrap out-of-bag (OOB)* y *bootstrap 0.632+* (12), si bien no se recogen en el presente trabajo.

## 5.2. Valores perdidos

Los valores perdidos se presentan de forma habitual en investigación biomédica. Los estudios retrospectivos suelen verse afectados en mayor medida por esta problemática, dado que emplean datos que han sido recolectados previamente; habitualmente por razones clínicas y no bajo un diseño de investigación que permita minimizar su aparición.

Los valores perdidos pueden producirse bajo diferentes mecanismos. Ésta es una cuestión fundamental a tener en cuenta al abordar su tratamiento, dado que las diferentes estrategias existentes para ello realizan asunciones determinadas en este sentido.

### 5.2.1. Mecanismos de pérdida

En relación con los mecanismos de pérdida de información, es habitual el empleo de la tipología establecida por Rubin (13). Esta distingue entre *missing completely at random* (MCAR), *missing at random* (MAR) y *missing not at random* (MNAR).

*Missing completely at random* (MCAR): los valores perdidos ocurren de forma completamente aleatoria, de manera que los individuos con valores perdidos son una muestra representativa de la población con datos completos. Ejemplos de este mecanismo son la pérdida de datos debida a errores administrativos o a fallos en el equipo necesario para registrar el resultado de una prueba diagnóstica.

*Missing at random* (MAR): la probabilidad de que un valor esté perdido se relaciona únicamente con las variables observadas en la muestra. Por ejemplo, en relación con el resultado de una prueba diagnóstica que requiere la colaboración del paciente, la probabilidad de que se tenga un valor perdido (por no haberse podido realizar) puede estar relacionada con la edad del paciente, encontrándose valores perdidos con mayor frecuencia en pacientes pediátricos o geriátricos.

*Missing not at random* (MNAR): la probabilidad de que un valor esté perdido depende del propio valor perdido, o bien de variables no recogidas en la muestra. Por ejemplo, en preguntas de un cuestionario relacionadas con el consumo de sustancias psicoactivas, los pacientes que consumen estas sustancias pueden no responder en mayor medida que aquellos que no las consumen.

### 5.2.2. Imputación múltiple

Las técnicas clásicas de modelización estadística requieren un tratamiento previo de los valores perdidos para poder ser aplicadas. Cuando los valores perdidos no suponen una pérdida de información importante en las variables relevantes en el análisis (p. e., el porcentaje de observaciones con valores perdidos es inferior al 5%), restringir el análisis a los individuos que presentan datos completos puede resultar adecuado; esta estrategia se conoce como *análisis de casos completos*. Sin embargo, si la cantidad de información perdida es mayor, el análisis de casos completos conlleva una disminución de potencia estadística y, en muchas situaciones, una estimación sesgada de los efectos en el modelo (14).

En estas circunstancias, puede resultar ventajoso el empleo de técnicas de imputación múltiple, en las que los valores perdidos se reemplazan por distintos valores plausibles, en función de los valores observados en otras variables. Debe tenerse en cuenta que las potenciales ventajas de la imputación múltiple frente al análisis de casos completos dependen de cuál sea el patrón de los valores perdidos y de las relaciones existentes entre las variables recogidas en la muestra, así como de que los métodos de imputación empleados sean adecuados. En los estudios de Lee et al. (14) y Hughes et al. (15) se proporciona un marco para el empleo de técnicas de imputación múltiple.

La imputación múltiple permite tratar valores perdidos de mecanismo MCAR, MAR y MNAR. No obstante, el empleo de la imputación múltiple bajo valores perdidos de mecanismo MNAR implica asunciones adicionales que afectan a las imputaciones (5).

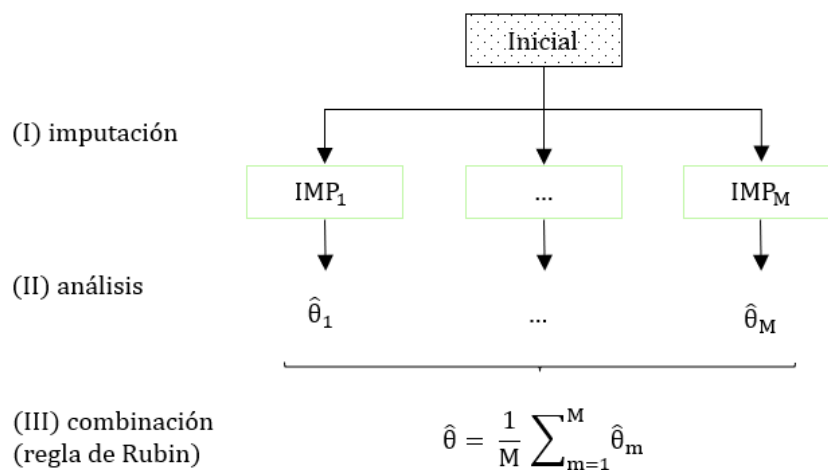
La aplicación de técnicas de imputación múltiple involucra tres etapas (Figura 2): (I) imputación, (II) análisis y (III) combinación (*pooling*) de las estimaciones obtenidas (5).

- (I) En primer lugar, cada uno de los valores perdidos se reemplaza por  $M$  valores plausibles ( $M > 1$ ). Bajo el marco teórico de la imputación múltiple, entre 3 y 5 imputaciones se consideran suficientes si la presencia de valores perdidos no es excesiva (16); sin embargo, siempre que la capacidad computacional no suponga una limitación, se recomienda aumentar este valor (17,18). Como resultado, se

tienen múltiples versiones completas de la muestra inicial, que difieren en los valores imputados. Estas diferencias recogen el grado de incertidumbre acerca de los valores perdidos.

- (II) El análisis estadístico se realiza sobre cada una de las muestras imputadas, siguiendo la misma metodología que se emplearía sobre una muestra inicialmente completa. En consecuencia, se obtienen  $M$  estimaciones para el parámetro de interés ( $\theta$ ):  $\hat{\theta}_1, \dots, \hat{\theta}_M$ .
- (III) Finalmente, a partir de las  $M$  estimaciones  $\hat{\theta}_1, \dots, \hat{\theta}_M$  se obtiene una estimación combinada (*pooled*), típicamente mediante la regla de Rubin (16):  $\hat{\theta} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}_m$ .

Naturalmente, se puede tener interés en un vector de parámetros (por ejemplo, el vector de coeficientes de un modelo de regresión); el procedimiento descrito es válido para el caso multidimensional.



**Figura 2.** Etapas del proceso de imputación múltiple. (I) imputación de los valores perdidos (representados mediante el fondo punteado); (II) realización del análisis estadístico sobre cada una de las  $M$  versiones imputadas de la muestra y obtención de  $M$  estimaciones del parámetro de interés  $\theta$ :  $\hat{\theta}_1, \dots, \hat{\theta}_M$ ; (III) obtención de una estimación combinada (*pooled*) de  $\theta$  mediante la regla de Rubin.

### 5.2.2.1. Métodos univariantes de imputación múltiple

El marco univariante en la imputación múltiple de valores perdidos se caracteriza por contar con una única variable incompleta a imputar, estando el resto de variables completamente observadas. Existen diversos métodos univariantes para la imputación múltiple de valores perdidos; a continuación, se describen los empleados en este trabajo: *predictive mean matching* y *regresión logística bayesiana*.

Se denota con  $Y = (Y_{\text{obs}}, Y_{\text{mis}})$  a la variable incompleta, donde  $Y_{\text{obs}}$  denota sus valores observados e  $Y_{\text{mis}}$  sus valores perdidos. El número de valores observados y perdidos en  $Y$  es, respectivamente,  $n_{\text{obs}}$  y  $n_{\text{mis}}$ , siendo  $n = (n_{\text{obs}}, n_{\text{mis}})$  el tamaño muestral. Por otra parte, se denota por  $X = (X_1, \dots, X_j)$  al conjunto de variables completamente observadas.  $X_{\text{obs}}$  denota los valores de estas variables que corresponden a  $Y_{\text{obs}}$  y  $X_{\text{mis}}$  los que corresponden a  $Y_{\text{mis}}$ .

#### 5.2.2.1.1. *Predictive mean matching*

El método *predictive mean matching* es ampliamente utilizado para la imputación de variables cuantitativas, continuas o semi-continuas. Se trata de un método de imputación de los denominados *hot-deck*, en los que los valores perdidos se reemplazan por valores observados en la misma variable, tomados de observaciones similares a aquella que presenta el valor faltante.

El método *predictive mean matching* es un método de imputación múltiple de enfoque bayesiano, en el que la imputación de los valores perdidos se realiza mediante el siguiente algoritmo (16,19):

1. Se modeliza  $Y_{\text{obs}}$  dado  $X_{\text{obs}}$  empleando una regresión lineal,  $Y_{\text{obs}} = X_{\text{obs}}^T \beta + e$  e  $e \sim N(0, \sigma^2 I)$ . Bajo el marco bayesiano, los parámetros del modelo son considerados variables aleatorias a las que se asigna una distribución *a priori*; en este contexto, suelen emplearse distribuciones *a priori* de carácter no informativo.

2. Se estiman  $\hat{\beta}$  y  $\hat{\sigma}$  por mínimos cuadrados ordinarios.
3. Se selecciona aleatoriamente  $\sigma^{2*}$ , procedente de la distribución a posteriori de  $\sigma^2$ . Bajo distribución a priori no informativa, la distribución *a posteriori* de  $\sigma^2$  es  $\frac{\hat{e}^T \hat{e}}{\chi_{\hat{n}_{obs}-r}^2}$ , siendo  $r$  el número de coeficientes del modelo de regresión lineal.
4. Se selecciona aleatoriamente  $\beta^*$ , procedente de la distribución conjunta a posteriori de  $\beta$ . Bajo distribución conjunta a priori no informativa, esta distribución es  $N(\hat{\beta}, V(\hat{\beta}))$ , donde  $V(\hat{\beta}) = \sigma^2 (X_{obs}^T X_{obs})^{-1}$ . Se toma  $\sigma^2 = \sigma^{2*}$ .
5. Se obtienen  $\hat{Y}_{obs} = X_{obs} \hat{\beta}$  e  $\hat{Y}_{mis} = X_{mis} \beta^*$ .
6. Para cada valor perdido  $\hat{Y}_{mis,i}$ , se determina el vector  $\Delta = |\hat{Y}_{obs} - \hat{Y}_{mis,i}|$ . Se selecciona aleatoriamente uno de los  $d$  menores valores de  $\Delta$  (habitualmente,  $d = 3$ ,  $d = 5$  o  $d = 10$ ) y se toma el correspondiente valor de  $Y_{obs}$  para imputar  $Y_{miss,i}$ .
7. El proceso se repite  $M$  veces, obteniéndose  $M$  muestras completas.

#### 5.2.2.1.2. Regresión logística bayesiana

La regresión logística bayesiana es un método de imputación múltiple para variables de naturaleza dicotómica, en el que la imputación de valores perdidos se lleva a cabo a través del algoritmo siguiente (16,19):

1. Se modeliza  $Y_{obs}$  dado  $X_{obs}$  empleando una regresión logística,  $\ln\left(\frac{p(Y_{obs}=1)}{p(Y_{obs}=0)}\right) = X_{obs}^T \beta$ .
2. Se estiman  $\hat{\beta}$  y  $\hat{\sigma}$  por máxima verosimilitud.

3. Se selecciona aleatoriamente  $\beta^*$ , procedente de la distribución conjunta a posteriori de  $\beta$ . Bajo distribución conjunta a priori no informativa, esta distribución se aproxima a una distribución  $N(\hat{\beta}, V(\hat{\beta}))$ .
4. Para cada valor perdido  $Y_{\text{mis},i}$ , se obtiene  $\hat{p}(Y_{\text{mis},i} = 1) = \frac{e^{x_i^T \beta^*}}{1 + e^{x_i^T \beta^*}}$ .
5. Cada valor perdido  $Y_{\text{mis},i}$  se imputa mediante un valor  $Y_{\text{mis},i}^*$  procedente de una distribución Bernouilli( $\hat{p}(Y_{\text{mis},i} = 1)$ ).
6. El proceso se repite  $M$  veces, obteniéndose  $M$  muestras completas.

#### 5.2.2.2. Imputación multivariante por ecuaciones encadenadas (MICE)

La imputación multivariante por ecuaciones encadenadas (*Multivariate Imputation by Chained Equations* o MICE) es un procedimiento de enfoque bayesiano desarrollado para abordar la imputación múltiple de valores perdidos cuando se tiene más de una variable incompleta a imputar.

En este escenario, la imputación de los valores perdidos se basa en la distribución multivariante  $P(Y | \theta)$  correspondiente a  $Y = (Y_1, \dots, Y_k)$ ; por sencillez en la notación, todas las variables se consideran incompletas. En MICE, la distribución  $P(Y | \theta)$  se define de forma implícita, mediante la especificación de las distribuciones condicionadas  $P(Y_j | Y_{-j}, \theta_j)$  correspondientes a cada  $Y_j$  (donde  $Y_{-j}$  denota a las variables en  $Y$  excluyendo  $Y_j$  y el vector de parámetros  $\theta_j$  es específico de la distribución condicionada correspondiente, con  $j = 1, \dots, k$ ) (20).

Cada distribución  $P(Y_j | Y_{-j}, \theta_j)$  se emplea para imputar  $Y_j^{\text{mis}}$  a partir de  $Y_{-j}$ . Por ejemplo, mediante los métodos univariantes de imputación anteriormente presentados, esta distribución se especifica través de una regresión lineal o de una regresión logística binomial, atendiendo a la naturaleza de la variable a imputar. En MICE, estos métodos univariantes se aplican de forma encadenada para la imputación de cada variable  $Y_j$ , tras establecer una

secuencia u orden de imputación. Así, los valores imputados en una variable pasan a ser considerados como observados para la imputación de las siguientes.

Se trata de un método iterativo, en el que se parte de unas imputaciones iniciales (por ejemplo, valores aleatoriamente tomados entre los observados para cada variable) y cada iteración corresponde a un ciclo completo sobre todas las variables a imputar. Los valores imputados en una determinada iteración son empleados para realizar las imputaciones en la iteración siguiente. De esta manera, una determinada iteración  $t$  consiste en la siguiente sucesión de muestras del muestreador de Gibbs (20):

$$\begin{aligned} \theta_1^{*(t)} &\sim P(\theta_1 \mid Y_1^{\text{obs}}, Y_2^{(t-1)}, \dots, Y_k^{(t-1)}) \\ Y_1^{*(t)} &\sim P(Y_1^{\text{mis}} \mid Y_1^{\text{obs}}, Y_2^{(t-1)}, \dots, Y_k^{(t-1)}, \theta_1^{*(t)}) \\ &\dots \\ \theta_k^{*(t)} &\sim P(\theta_k \mid Y_k^{\text{obs}}, Y_1^{(t)}, \dots, Y_{k-1}^{(t)}) \\ Y_k^{*(t)} &\sim P(Y_k^{\text{mis}} \mid Y_k^{\text{obs}}, Y_1^{(t)}, \dots, Y_{k-1}^{(t)}, \theta_k^{*(t)}) \end{aligned}$$

Finalmente, los valores perdidos se reemplazan por las imputaciones obtenidas en la última iteración. Se suele trabajar con un número reducido de iteraciones, generalmente 5 o 10 se consideran suficientes.

Este proceso iterativo se lleva a cabo  $M$  veces, para obtener  $M$  muestras completas.

### 5.2.2.3. Valores perdidos en la variable de interés en el modelo pronóstico: enfoques MI y MID

Cuando la propia variable de interés en el modelo pronóstico presenta valores perdidos, la aplicación de las técnicas de imputación múltiple puede llevarse a cabo siguiendo distintos enfoques:

- (I) Enfoque MI (*Multiple Imputation*): las observaciones con valores perdidos en la variable de interés  $Y$  se emplean tanto en la fase de imputación como en la fase de análisis.
- (II) Enfoque MID (*Multiple Imputation, then Deletion*): las observaciones con valores perdidos en la variable de interés  $Y$  se emplean en la fase de imputación, pero se descartan con anterioridad a la realización del análisis estadístico.

Von Hippel (22) estudió estos enfoques de forma teórica y mediante simulación. Las conclusiones de su investigación indican que el enfoque MID es, generalmente, preferible: bajo este enfoque, la estimación de los coeficientes de regresión se ve protegida frente a valores imputados de  $Y$  que puedan resultar problemáticos o no adecuados; y los estimadores presentan mayor precisión. No obstante, en escenarios determinados en los que las observaciones con valores perdidos en  $Y$  recogen información adicional sustancial acerca de la regresión de interés, el enfoque MI puede resultar ventajoso.

### 5.3. Integración de la imputación múltiple en la validación bootstrap

La forma en que la imputación múltiple debe integrarse en el marco de la validación bootstrap se encuentra actualmente en estudio.

Para la obtención del rendimiento aparente del modelo pronóstico, la incorporación de la imputación múltiple resulta clara:

- (I) En primer lugar, se imputan los valores perdidos de la muestra, obteniéndose  $M$  muestras completas.
- (II) En cada una de ellas, se desarrolla el modelo pronóstico y se obtiene su rendimiento aparente, que resulta de su evaluación en la misma muestra imputada con que ha sido desarrollado.

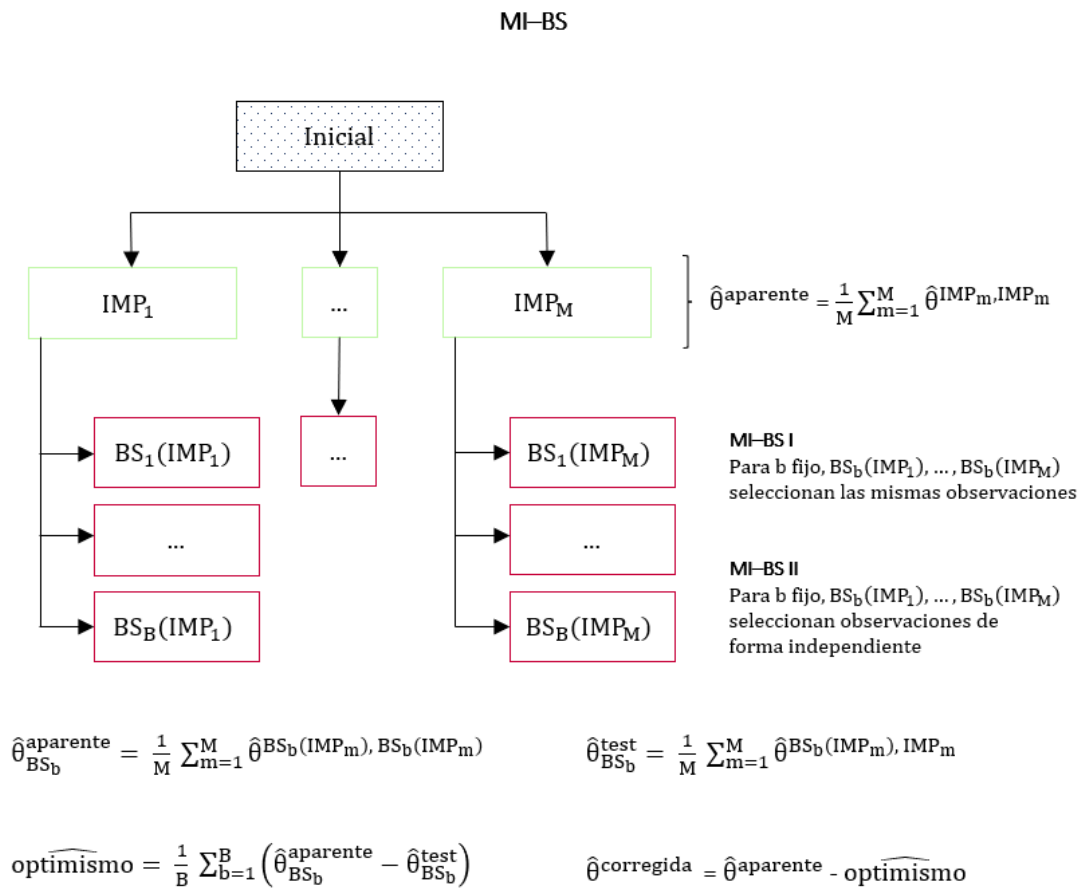
- (III) La aplicación de la regla de Rubin ofrece la medida combinada de rendimiento aparente del modelo:  $\hat{\theta}^{\text{aparente}} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}^{\text{aparente}_m}$ .

Sin embargo, para la estimación del optimismo, la imputación múltiple puede combinarse con el remuestreo bootstrap siguiendo diferentes estrategias. Las estrategias consideradas en este trabajo, que se denominarán MI-BS y BS-MI, son descritas por Musoro et al. en (6). Estas estrategias se detallan a continuación.

### 5.3.1. Estrategia MI-BS

La estrategia MI-BS se caracteriza por llevar a cabo, en primer lugar, la imputación múltiple de valores perdidos sobre la muestra inicial; y, a continuación, realizar el remuestreo bootstrap sobre cada una de las  $M$  muestras completas. El procedimiento presenta dos posibles variaciones (Figura 3):

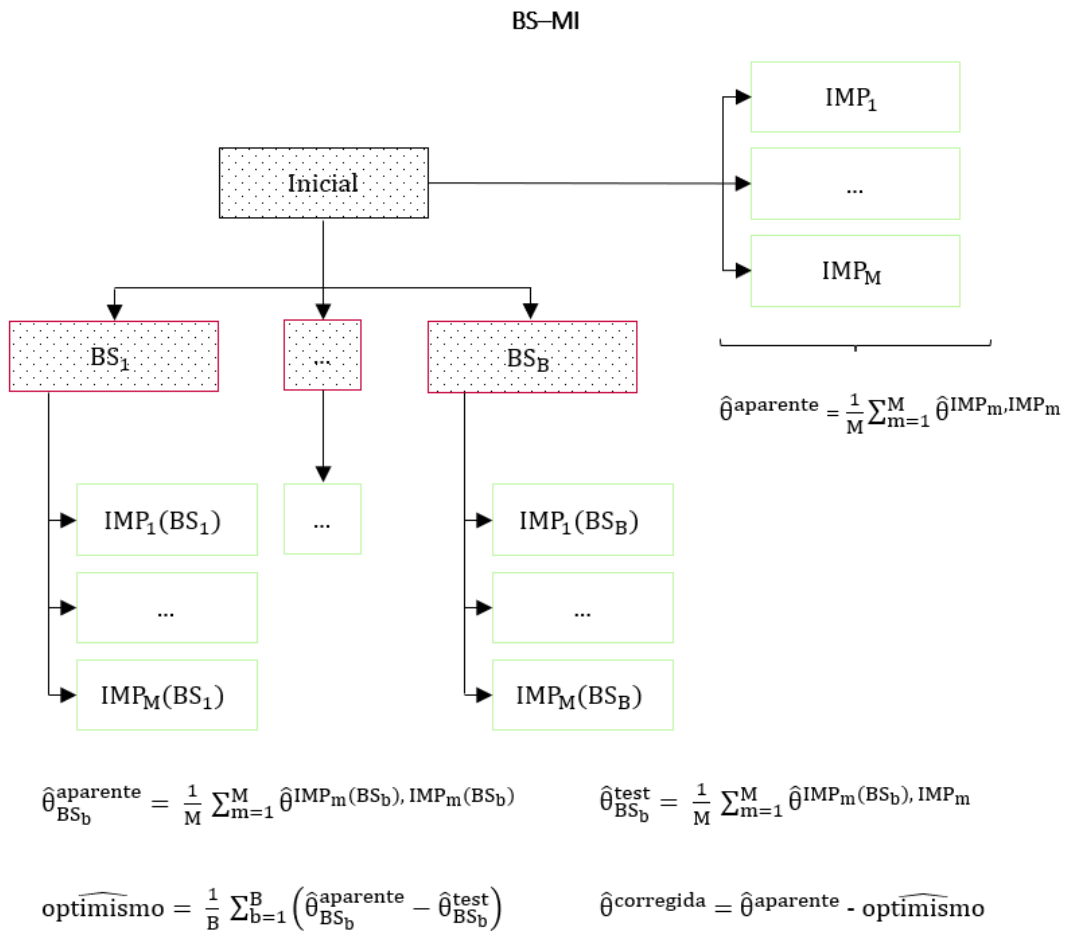
- (I) MI-BS 1: el remuestreo bootstrap se realiza de forma conjunta sobre las  $M$  muestras completas, seleccionando unos mismos individuos en todas ellas.
- (II) MI-BS 2: el remuestreo bootstrap se realiza de manera independiente en cada una de las  $M$  muestras completas.



**Figura 3.** Estrategia MI-BS. El fondo punteado indica la presencia de valores de valores perdidos.  $\theta$  denota a la medida de rendimiento de interés; el primer superíndice denota la muestra empleada para el desarrollo del modelo pronóstico y el segundo superíndice denota la muestra sobre la que el modelo se evalúa.

### 5.3.2. Estrategia BS-MI

En la estrategia BS-MI, el remuestreo bootstrap se realiza directamente sobre la muestra inicial con valores perdidos; por lo tanto, las  $B$  muestras bootstrap pueden presentar también valores perdidos, en función de cuáles sean los individuos seleccionados en cada una de ellas. La imputación múltiple de valores perdidos se lleva a cabo, posteriormente, sobre cada una de estas  $B$  muestras bootstrap (Figura 4).



**Figura 4.** Estrategia BS-MI. El fondo punteado indica la presencia de valores perdidos.  $\theta$  denota a la medida de rendimiento de interés; el primer superíndice denota la muestra empleada para el desarrollo del modelo pronóstico y el segundo superíndice denota la muestra sobre la que el modelo se evalúa.

#### 5.4. Regresión logística binomial

En investigación biomédica, con frecuencia la variable de interés corresponde a que un paciente presente o no un determinado evento (por ejemplo, una determinada patología) y es, por lo tanto, dicotómica o binaria. Ante variables de esta naturaleza, es habitual el empleo de modelos de regresión logística binomial.

La regresión logística binomial permite modelizar la probabilidad de ocurrencia del evento de interés condicionada a una serie de covariables. Se trata de un modelo lineal generalizado, en

el que la función de enlace es la función *logit*, el logaritmo neperiano del *odds*. Las dos posibles expresiones matemáticas del modelo, equivalentes, son las que siguen:

$$\ln \left( \frac{p(Y_i = 1)}{p(Y_i = 0)} \right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} \quad i = 1, \dots, n$$

$$p(Y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})}} \quad i = 1, \dots, n$$

donde  $\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$  recibe la denominación de *predictor lineal*.

## 5.5. Medidas de rendimiento predictivo

El rendimiento predictivo de un modelo pronóstico involucra los aspectos de discriminación y calibración. A continuación, se describen algunas de las medidas de rendimiento empleadas habitualmente en la validación de modelos logísticos binomiales.

### 5.5.1. Discriminación

La discriminación refiere a la capacidad del modelo pronóstico para distinguir entre aquellos individuos que presentan el evento de interés y aquellos que no lo hacen, asignando probabilidades de ocurrencia más elevadas a los primeros y más reducidas a los segundos (23).

Existen diversas medidas para evaluar la capacidad de discriminación de un modelo pronóstico. En modelos lineales generalizados, una medida de discriminación habitualmente empleada es el *estadístico de concordancia* o *estadístico C* (2).

### 5.5.1.1. Estadístico C

El estadístico C es un estadístico basado en rangos, relacionado con la D de Somers (24). Para variables de naturaleza dicotómica o binaria, el estadístico C considera todos los pares conformados por un individuo que presenta el evento de interés y otro que no; y se obtiene como la proporción de estos pares en los que la probabilidad de ocurrencia predicha por el modelo es superior para el primero de ellos (los empates se contabilizan parcialmente).

$$\hat{C} = \frac{n_{\text{concordantes}} + 0.5 n_{\text{empates}}}{n}$$

donde  $n_{\text{concordantes}}$  denota el número de pares en los que el modelo asigna una probabilidad de ocurrencia mayor al individuo que presenta el evento,  $n_{\text{empates}}$  denota el número de pares en los que el modelo asigna una misma probabilidad a ambos individuos y  $n$  denota al número total de pares formados por un individuo que presenta el evento y otro que no lo hace.

En este caso, el estadístico C coincide con el área bajo la curva ROC; estimando la probabilidad de que, seleccionados al azar un individuo con el evento de interés y un individuo sin él, el modelo asigne una probabilidad de ocurrencia más elevada al primero de los dos.

### 5.5.2. Calibración

La buena discriminación de un modelo pronóstico no garantiza que las predicciones ofrecidas por este sean adecuadas. Por ejemplo, un modelo pronóstico puede discriminar satisfactoriamente entre aquellos individuos que presentan el evento y aquellos que no; y, sin embargo, ofrecer predicciones para la probabilidad de ocurrencia que sean, de manera sistemática, demasiado elevadas para todos los individuos.

La precisión de las predicciones ofrecidas por el modelo, entendida como el grado de acuerdo entre estas y el número de eventos observados, se denomina calibración (25). La calibración de un modelo pronóstico puede ser evaluada mediante las medidas *calibration-in-the-large* (CITL) y *calibration slope*.

### 5.5.2.1. *Calibration-in-the-large* (CITL)

La medida *calibration-in-the-large* (CITL) se relaciona con la diferencia entre la predicción media para la probabilidad de ocurrencia ofrecida por el modelo pronóstico y la prevalencia del evento. Así, un modelo pronóstico bien calibrado presenta un valor de CITL próximo a cero cuando se evalúa en nuevos individuos. Valores de CITL estrictamente positivos indican que el modelo pronóstico ofrece predicciones para la probabilidad de evento que son, en media, excesivamente elevadas; mientras que valores de CITL estrictamente negativos indican que estas predicciones son, en media, excesivamente reducidas.

Para modelos lineales generalizados, en la misma muestra empleada para el desarrollo del modelo pronóstico la media de las predicciones coincide con la proporción de eventos y la CITL toma el valor cero; esto se garantiza por la inclusión en el modelo de un término independiente. Sin embargo, esta correspondencia puede no darse cuando el modelo se evalúa sobre una muestra diferente.

Para modelos logísticos binomiales, la estimación de la CITL corresponde a la estimación de  $\alpha$  en el siguiente modelo (2):

$$\ln\left(\frac{p(Y_i = 1)}{p(Y_i = 0)}\right) = \alpha + PL_i \quad i = 1, \dots, n$$

donde  $PL_i$  es el predictor lineal del modelo pronóstico, que se introduce como única covariable y asociado a un efecto igual a 1 (*offset*).

Es importante notar que un valor próximo a cero para CITL no garantiza una buena calibración del modelo pronóstico, dado que puede ocurrir que la infrapredicción en determinados individuos se vea compensada por la sobrepredicción en otros (26).

### 5.5.2.2. *Calibration slope*

Otra medida de calibración empleada en modelos pronósticos es la pendiente de calibración (*calibration slope*). La pendiente de calibración está relacionada con la dispersión de las predicciones ofrecidas por el modelo pronóstico, e idealmente debe tomar el valor uno. Valores inferiores a uno se corresponden con una dispersión excesiva: el modelo proporciona predicciones para la probabilidad de evento que son excesivamente elevadas para individuos de alto riesgo y excesivamente reducidas para individuos de bajo riesgo (27); por el contrario, valores superiores a uno indican que las predicciones dadas por el modelo no varían lo suficiente, encontrándose en un rango demasiado estrecho.

Sobre la misma muestra en que se desarrolla el modelo pronóstico, bajo los métodos de estimación de mínimos cuadrados y de máxima verosimilitud, la pendiente de calibración toma el valor uno. Sin embargo, en la validación interna de modelos pronósticos, la pendiente de calibración presenta habitualmente valores inferiores, indicativos de sobreajuste (27). En estas circunstancias, la pendiente de calibración indica cómo deben modificarse los coeficientes estimados del modelo para proporcionar predicciones mejor calibradas en nuevos individuos de la misma población (28). En este sentido, su estimación se emplea como factor de contracción (*shrinkage factor*) en el proceso de recalibración del modelo pronóstico, por lo que la pendiente de calibración se denomina también de esta manera.

Para modelos logísticos binomiales, la estimación de la pendiente de calibración corresponde a la estimación de  $\beta$  en el modelo que sigue:

$$\ln \left( \frac{p(Y_i = 1)}{p(Y_i = 0)} \right) = \alpha + \beta PL_i \quad i = 1, \dots, n$$

donde  $PL_i$  es el predictor lineal del modelo pronóstico, que se introduce como única covariable.

Debe notarse que una pendiente de calibración de valor próximo a uno puede encontrarse también en modelos de calibración pobre; por este motivo, la calibración de un modelo

pronóstico debe evaluarse conjuntamente a través de la pendiente de calibración y de la CITL (26,27).

## 6. Planteamiento del problema: estudio de simulación

Para evaluar las diferentes estrategias de incorporación de la imputación múltiple de valores perdidos en la validación bootstrap, se lleva a cabo un estudio de simulación para una regresión logística binomial; considerando cinco predictores continuos con distribución normal multivariante y una variable de interés de naturaleza dicotómica.

El diseño general del estudio de simulación se realiza tomando como guía el estudio realizado por Steyerberg et al. (3), en el que se evalúan diferentes técnicas de validación interna de modelos pronósticos.

La simulación de los datos se lleva a cabo en dos fases: (I) generación de muestras completas y (II) introducción de valores perdidos. Una vez simulados los datos, se aplican los métodos descritos para la imputación de los valores perdidos, el desarrollo del modelo pronóstico y la obtención de las estimaciones de rendimiento predictivo a través de cada una de las estrategias de validación a comparar.

Se consideran distintos escenarios de simulación, bajo cada uno de los cuales se realizan  $NSIM = 500$  simulaciones.

El diseño del estudio de simulación se representa en la Figura 5 y se detalla en las secciones siguientes.

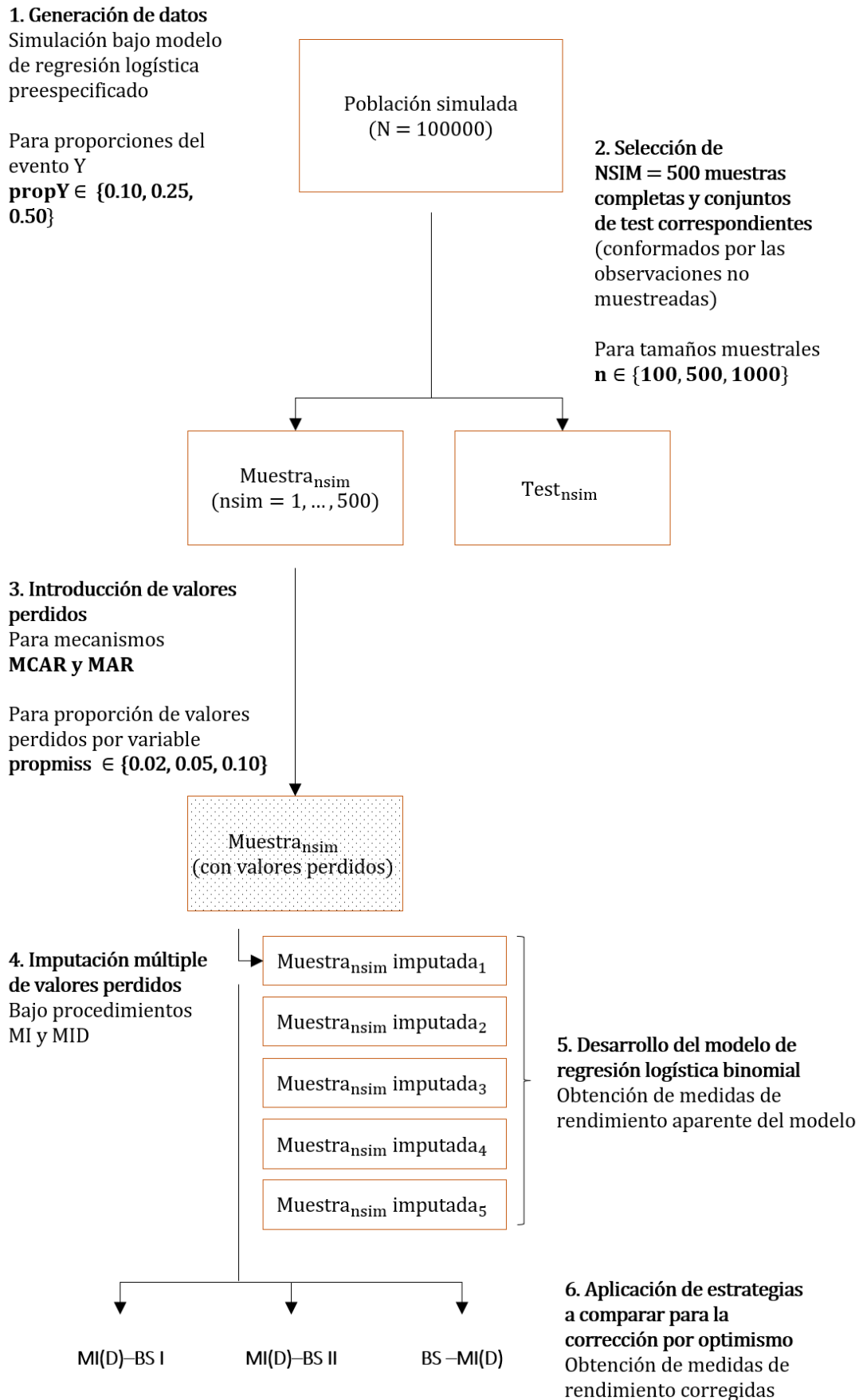


Figura 5. Diseño del estudio de simulación

## 6.1. Descripción de los escenarios del estudio

Con el objetivo de cubrir un amplio abanico de situaciones encontradas en la práctica en el marco del desarrollo de modelos pronósticos, se consideran diferentes escenarios de simulación. Estos presentan variaciones en tamaño muestral, proporción muestral de eventos, y proporción de valores perdidos por variable. En total, se contemplan 27 escenarios de simulación, bajo cada uno de los cuales se consideran valores perdidos de mecanismo MCAR y valores perdidos de mecanismo MAR (Tabla 1).

**Tabla 1.** Escenarios de simulación

Parámetro	Notación	Valores
Tamaño muestral	$n$	100, 500, 1000
Proporción muestral de eventos	$propY$	0.10, 0.25, 0.50
Proporción de valores perdidos por variable	$propmiss$	0.02, 0.05, 0.10
Mecanismo de pérdida	-	MCAR, MAR

La combinación de los valores considerados para el tamaño muestral y para la proporción muestral de eventos corresponde a valores para el número de eventos por variable (EPV) de 2, 5, 10, 20, 25, 50 y 100.

## 6.2. Generación de los datos

Los datos se generan bajo una regresión logística binomial, con cinco predictores continuos y una variable de interés de naturaleza dicotómica. Se generan tres conjuntos de datos de tamaño  $N = 100000$ , que difieren en la prevalencia del evento de interés. Para dicha prevalencia, se consideran los valores  $propY \in \{0.10, 0.25, 0.50\}$ .

Estos conjuntos de datos representan, en cada caso, la población de procedencia sobre la que, a continuación, se seleccionan las muestras para el desarrollo de los modelos pronósticos. Sobre estas muestras completas se introducen, posteriormente, los valores perdidos.

El procedimiento seguido para la generación de los datos es el siguiente:

1. Se simulan los valores correspondientes a los predictores continuos  $X = (X_1, \dots, X_5)$ , obteniéndose  $n$  vectores de datos procedentes de una distribución Normal multivariante de vector de medias  $\mathbf{0}$  y matriz de varianzas-covarianzas  $\Sigma$  con  $\text{Var}(X_j) = 1$  y  $\text{Cov}(X_j, X_{j'}) = 0.10$  para  $j \neq j'$ , con  $j = 1, \dots, 5$ ,  $j' = 1, \dots, 5$ ; el empleo de este último valor, distinto de cero, obedece a que el uso de técnicas de imputación múltiple tiene sentido en presencia de relaciones entre las variables que puedan ser aprovechadas en el proceso de imputación.
2. La probabilidad de evento se modeliza mediante una regresión logística binomial:

$$p(Y_i = 1 | X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_5 X_{5i})}} \quad i = 1, \dots, n$$

Se emplean como coeficientes de regresión  $\beta_1 = \beta_2 = 0.50$ ,  $\beta_3 = 0.40$  y  $\beta_4 = \beta_5 = 0.30$ . Estos coeficientes se establecen empíricamente, buscando que los modelos pronósticos presenten una capacidad de discriminación moderada cuando se desarrollan en condiciones óptimas (Anexo 1. Estadístico C de referencia). El término independiente  $\beta_0$  se establece resolviendo la ecuación  $\frac{1}{n} \sum_{i=1}^n p(Y_i = 1) = \text{prop}Y$  mediante métodos numéricos.

3. Para cada observación  $i$ , se genera  $Y_i$  como procedente de una distribución Bernoulli de parámetro  $p_i = p(Y_i = 1 | X)$ .

### 6.2.1. Obtención de muestras completas

Sobre cada uno de los tres conjuntos de datos generados de esta manera, se seleccionan  $NSIM = 500$  muestras de tamaño  $n \in \{100, 500, 1000\}$ . Con la finalidad de controlar la proporción muestral de eventos, se emplea un muestreo aleatorio estratificado en función de la variable dicotómica  $Y$ .

## 6.2.2. Introducción de valores perdidos

En cada muestra seleccionada, se introducen valores perdidos tanto en los predictores  $X$  como en la variable de interés  $Y$ . Ello responde (I) a que ésta es la situación encontrada habitualmente en la práctica; y (II) a que se busca cubrir en el presente trabajo una situación en la que la incorporación de la imputación múltiple en el marco de la validación bootstrap no ha sido estudiada.

Los valores perdidos se introducen mediante mecanismo MCAR y mediante mecanismo MAR, en una proporción por variable  $\text{propmiss} \in \{0.02, 0.05, 0.10\}$ . Debe notarse que la introducción de valores perdidos en la variable de interés  $Y$  puede modificar la proporción muestral de eventos  $\text{prop}Y$  y, en consecuencia, el número de eventos por variable que se tenía en la muestra completa; los valores indicados para la proporción muestral de eventos y para el EPV deben entenderse como aproximados.

### 6.2.2.1. Mecanismo MCAR

Los valores perdidos de mecanismo MCAR se introducen empleando, para cada observación de cada variable, una distribución Bernouilli de parámetro  $\text{propmiss}$  que determina la pérdida o no del valor correspondiente.

### 6.2.2.2. Mecanismo MAR

Los valores perdidos de mecanismo MAR se introducen bajo un enfoque similar al empleado por Marshall et al. (29) y por Wahl et al. (7), modelizando la probabilidad de pérdida a través de una regresión logística binomial.

Para cada uno de los predictores continuos  $X_j$ ,  $j = 1, \dots, 5$ , la probabilidad de pérdida se modeliza de la siguiente manera:

$$p(M_{X_{ji}} = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 M_{X_{li}} + \beta_2 X_{mi} \text{Iobs}_{X_{mi}} + \beta_3 Y_i \text{Iobs}_{Y_i})}} \quad l \neq m \neq j \quad i = 1, \dots, n$$

donde  $M_{X_{ji}}$  indica si el predictor continuo  $X_j$  presenta un valor perdido en la observación  $i$ ;  $X_{mi}$  denota el valor del predictor continuo  $X_m$  en la observación  $i$  e  $Iobs_{X_{mi}}$  indica si dicho valor está observado (es decir, no perdido);  $Y_i$  denota el valor de la variable  $Y$  en la observación  $i$  e  $Iobs_{Y_i}$  indica si dicho valor está observado; los predictores  $X_1$  y  $X_m$  son seleccionados aleatoriamente y sin reemplazamiento.

Se emplean como coeficientes de regresión  $\beta_1 = 1$  y  $\beta_2 = \beta_3 = 2$ ; los valores de estos coeficientes no revisten especial interés, el objetivo de la modelización es recoger la existencia de una relación entre la pérdida de valores en una determinada variable y los valores observados en otras. El término independiente  $\beta_0$  se establece resolviendo la ecuación  $\frac{1}{n} \sum_{i=1}^n p(M_{X_{ji}} = 1) = \text{propmiss}$  mediante métodos numéricos.

Para la variable  $Y$ , la probabilidad de pérdida se modeliza de manera similar:

$$p(M_{Y_i} = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} Iobs_{X_{1i}})}} \quad i = 1, \dots, n$$

donde  $M_{Y_i}$  indica si el valor de  $Y$  en la observación  $i$  está perdido;  $X_{1i}$  denota el valor del predictor  $X_1$  en la observación  $i$  e  $Iobs_{X_{1i}}$  indica si el valor de dicho predictor en la observación  $i$  está observado.

Se emplea como coeficiente de regresión  $\beta_1 = 2$ . El término independiente  $\beta_0$  se establece resolviendo la ecuación  $\frac{1}{n} \sum_{i=1}^n p(M_{Y_i} = 1) = \text{propmiss}$  mediante métodos numéricos.

Finalmente, la introducción o no del valor perdido en cada observación de cada variable se determina empleando una distribución Bernouilli de parámetro  $p(M_{X_{ji}} = 1)$  o  $p(M_{Y_i} = 1)$ , según corresponda.

### 6.3. Imputación múltiple

Los valores perdidos introducidos en cada muestra se imputan  $M = 5$  veces a través del procedimiento MICE con 5 iteraciones.

Para la imputación de la variable dicotómica  $Y$ , se emplea el método de regresión logística bayesiana; para la imputación de los predictores continuos  $X$ , se emplea el método *predictive mean matching*. Todas las variables se consideran como covariables en los modelos de imputación; en particular, atendiendo a la recomendación de Moons et al. (21), la variable  $Y$  se introduce como covariable en los modelos de imputación de los predictores  $X$ .

El procedimiento de imputación múltiple se lleva a cabo siguiendo dos enfoques distintos ante la presencia de valores perdidos en la variable de interés  $Y$ : MI (*Multiple Imputation*) y MID (*Multiple Imputation, then Deletion*); los resultados del estudio de simulación se presentan para cada uno de ellos.

### 6.4. Modelización y medidas de rendimiento predictivo

En cada muestra, tras la imputación múltiple de los valores perdidos, se desarrolla un modelo de regresión logística binomial para la variable dicotómica  $Y$ , considerando como covariables los cinco predictores continuos. No se consideran métodos de selección de predictores.

Como medidas de rendimiento predictivo, se consideran medidas de discriminación (estadístico C) y de calibración (CITL y *calibration slope*).

### 6.5. Estrategias a comparar

Las estrategias a comparar son las presentadas en la sección 5.3, denotadas MI-BS (distinguiendo las variantes MI-BS 1 y MI-BS 2) y BS-MI. Bajo el enfoque MID para la imputación múltiple, estas estrategias se denotan MID-BS y BS-MID. En todas las estrategias, el número de muestras bootstrap generadas es  $B = 100$ .

Musoro et al. (6) y Wahl et al. (7) consideran, dentro del bloque de estrategias MI-BS, una variante caracterizada por que la variable  $Y$  no se considera para la imputación de los predictores  $X$ . Dicha variante no se contempla en el presente trabajo, de acuerdo con los resultados del estudio realizado por Moons et al. (21), que concluye que esta decisión resulta desventajosa para el desarrollo de los modelos de regresión.

## 6.6. Evaluación de las estrategias

Para cada muestra, se considera como conjunto de test el conformado por todas aquellas observaciones procedentes del mismo conjunto inicial de  $N = 100000$  datos, excluyendo las observaciones muestreadas. Así, por ejemplo, para una muestra de tamaño  $n = 100$ , el conjunto de test lo conforman las 99900 observaciones restantes del conjunto de datos correspondiente que desempeña el rol poblacional.

En cada muestra, para el modelo de regresión logística binomial desarrollado, se obtienen:

- (I) Las medidas de rendimiento (estadístico  $C$ , CITL y *calibration slope*) sobre el conjunto de test, consideradas de referencia o *gold standard*. Para la obtención de estas medidas de referencia, se emplea el modelo de regresión logística binomial de coeficientes combinados (*pooled*), obtenido tras llevar a cabo el proceso de imputación múltiple.
- (II) Las estimaciones de estas medidas de rendimiento a través de cada una de las estrategias de validación a comparar: MI(D)-BS 1, MI(D)-BS 2 y BS-MI(D).
- (III) La diferencia  $\Delta$  entre la estimación de rendimiento y el correspondiente valor de referencia, para cada medida de rendimiento y estrategia de validación. Un valor  $\Delta > 0$  indica una sobrestimación de la medida de rendimiento, mientras que un valor  $\Delta < 0$  indica una infraestimación.

Bajo cada escenario de simulación, se tienen  $NSIM = 500$  muestras; por lo tanto, se desarrollan 500 modelos pronósticos y se obtienen, para cada medida de rendimiento y estrategia de validación, 500 diferencias  $\Delta$ .

Las diferentes estrategias se evalúan en base a las distribuciones empíricas de estas diferencias, para cada una de las medidas de rendimiento consideradas. Es deseable una estrategia para la cual estas distribuciones presenten un centro próximo a cero; es decir, una estrategia que no tienda ni a sobrestimar ni a infraestimar el rendimiento del modelo pronóstico. Además, resulta deseable que la variabilidad de estas distribuciones sea lo más reducida posible.

Para la evaluación, se emplean gráficos de caja, en los que se representa la mediana, el primer cuartil ( $Q1$ ) y el tercer cuartil ( $Q3$ ) de la distribución; la altura de la caja corresponde al rango intercuartílico ( $IQR = Q3 - Q1$ ) y constituye una medida de variabilidad.

La información recogida en los gráficos de caja se complementa con la tabulación de los valores promedio  $\bar{\Delta}$  de las distribuciones, tomados como métrica de centralidad. Se reportan, asimismo, estadísticos de dispersión (desviación típica) y de posición (percentiles 2.5% y 97.5%, denotados respectivamente  $P2.5$  y  $P97.5$ , además de los cuartiles presentados en los gráficos de caja).

## 6.7. Software

El estudio de simulación se ha realizado utilizando R (versión 4.1.1). El código empleado para la simulación se presenta en el Anexo 19. Código R para la simulación.

La simulación de los datos ha requerido el uso de la librería *MASS*. La función *uniroot* de la librería *stats* se ha empleado para la determinación de  $\beta_0$  en los modelos logísticos empleados para la generación inicial de los datos y para la introducción de valores perdidos de mecanismo MAR.

La aplicación del procedimiento MICE para la imputación múltiple de valores perdidos se ha realizado con la librería *mice*. La validación interna se ha llevado a cabo empleando código propio. La librería *pROC* se ha utilizado para la obtención de medidas de rendimiento predictivo.

Los tiempos computacionales se han registrado a través de la función *Sys.time* de la librería *base*.

La librería *parallel* se ha utilizado para el procesamiento en paralelo.

## 7. Resultados

A continuación, se presentan los resultados del estudio de simulación realizado. En la sección 7.1, se recogen los resultados correspondientes al rendimiento predictivo y a su estimación mediante las estrategias de validación evaluadas; en la sección 7.2, se muestran los resultados correspondientes a los tiempos de ejecución requeridos para la aplicación de estas.

### 7.1. Rendimiento predictivo

Los resultados se presentan, de manera separada, para valores perdidos de mecanismo MCAR y para valores perdidos de mecanismo MAR. A efectos comparativos, se presentan también los resultados de rendimiento obtenidos en ausencia de valores perdidos; es decir, cuando se cuenta con la muestra completa para el desarrollo y la validación bootstrap de los modelos pronósticos.

#### 7.1.1. Rendimiento de referencia

El rendimiento de referencia de cada modelo pronóstico desarrollado es el que se obtiene cuando se emplea sobre su correspondiente conjunto de test, representando el rendimiento predictivo sobre nuevos individuos. Este rendimiento es el que se pretende estimar a través de la validación interna.

La simulación de los datos se ha realizado de manera que los modelos presenten una capacidad de discriminación moderada en nuevos individuos cuando se desarrollan en condiciones óptimas, con valores para el estadístico C en torno a 0.75 (Anexo 1. Estadístico C de referencia). Estas condiciones corresponden a que se cuente con un tamaño muestral y con una proporción muestral de eventos suficientes en relación con la complejidad del modelo a desarrollar; es decir, corresponden a un número suficiente de eventos por variable (EPV). En estos escenarios de simulación, se obtienen valores en torno a uno para la pendiente de calibración y en torno a cero para la CITL; es decir, valores próximos a los valores idóneos para estas medidas de calibración (Anexo 2. *Calibration slope* de referencia, Anexo 3. *Calibration-in-the-large* de referencia).

Cuando los modelos se desarrollan a partir de muestras de menor tamaño y cuando la proporción muestral de eventos es más reducida (es decir, cuando se cuenta con un menor EPV), se observa que la discriminación y la calibración de los modelos resultan más pobres. En estos escenarios, el estadístico C presenta valores inferiores a 0.75; la pendiente de calibración presenta valores inferiores a uno, indicativos de predicciones extremas en nuevos individuos; y la CITL presenta valores que se alejan en mayor medida del cero, tanto por exceso como por defecto.

### 7.1.2. Rendimiento aparente

Los resultados de rendimiento aparente muestran generalmente un exceso de optimismo en los escenarios de menor EPV. En los escenarios más desfavorables, los valores aparentes del estadístico C exceden a los valores de referencia, en mediana, en torno a 0.10 (Anexo 4. Diferencia  $\Delta$  estadístico C sin validación interna). Por otra parte, los valores aparentes de la pendiente de calibración se sitúan por encima de los valores de referencia; por lo tanto, más próximos al valor idóneo que estos últimos (Anexo 5. Diferencia  $\Delta$  *calibration slope* sin validación interna). En cuanto a la CITL, las diferencias entre los valores aparentes y los valores de referencia se acrecientan en estos escenarios, presentando ambos signos (Anexo 6. Diferencia  $\Delta$  CITL sin validación interna). Estos resultados reflejan el fenómeno de sobreajuste en el desarrollo de modelos pronósticos y la necesidad de validación interna de los mismos.

### 7.1.3. Estimación del rendimiento a través de las estrategias de validación

A continuación, se recogen los resultados correspondientes a las diferentes estrategias de validación a comparar. En favor de la brevedad, en esta sección se presentan únicamente los resultados obtenidos bajo valores perdidos de mecanismo MAR, por tratarse de un mecanismo generalmente más plausible en la práctica. En Anexos pueden consultarse los resultados obtenidos bajo valores perdidos de mecanismo MCAR.

Asimismo, en Anexos se reportan estadísticos de centralidad, dispersión y posición correspondientes a las distribuciones empíricas de las diferencias  $\Delta$ , acompañados del número de simulaciones consideradas en cada caso para la obtención de dichos estadísticos (NSIM).

Los valores  $NSIM < 500$  se deben a la depuración de estimaciones obtenidas bajo falta de convergencia.

#### 7.1.3.1. Discriminación

Los resultados correspondientes al estadístico C permiten observar la existencia de diferencias en su estimación en función de cuál sea la estrategia empleada (Figura 6).

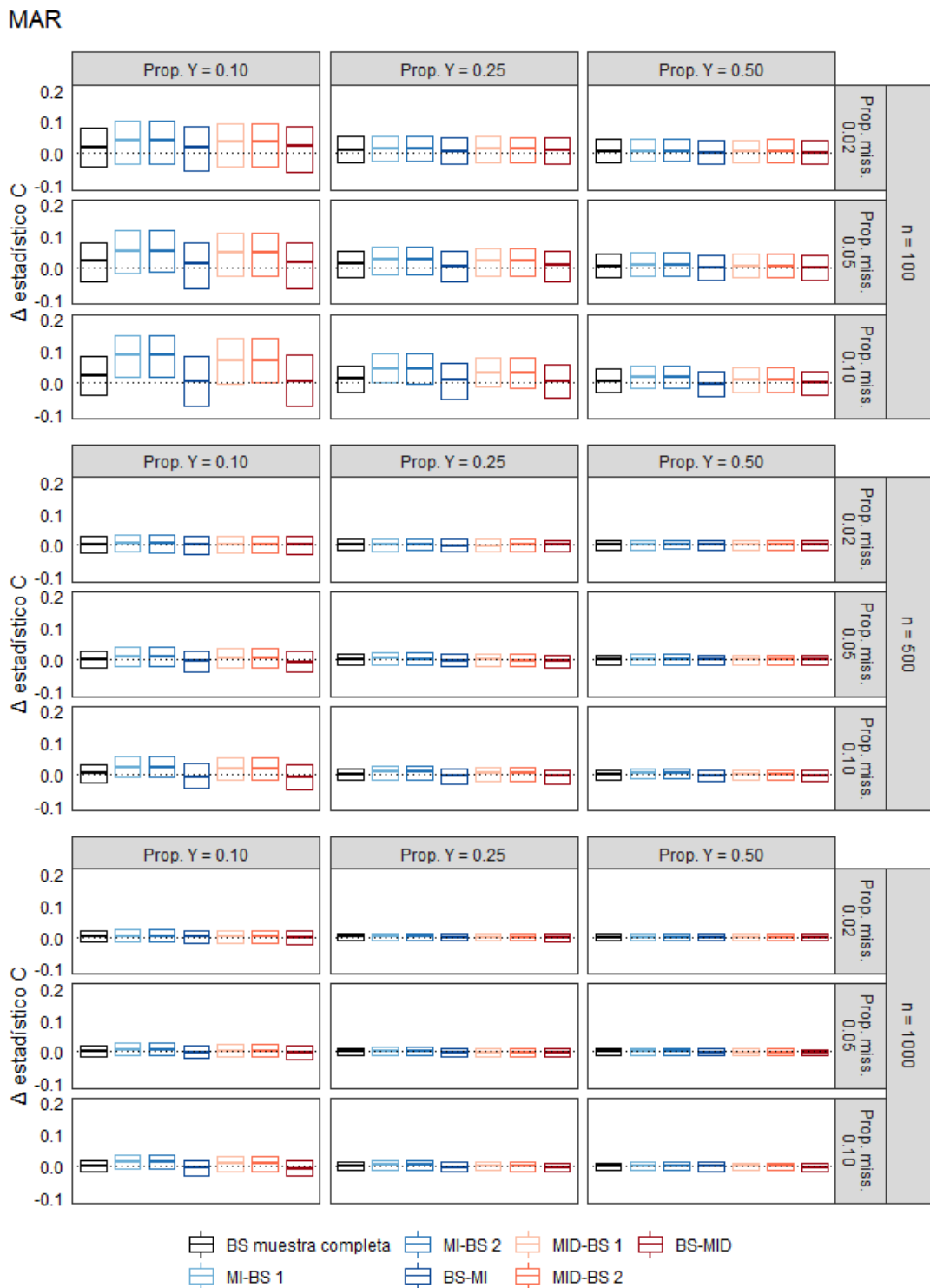
En escenarios con menor número de eventos por variable (EPV), las estrategias BS-MI(D) ofrecen estimaciones más realistas del estadístico C; es decir, más próximas a los valores obtenidos cuando los modelos desarrollados se evalúan sobre nuevas observaciones. Las estrategias MI(D)-BS, por el contrario, tienden a sobrestimar en mayor medida la discriminación de los modelos pronósticos, sin que se aprecien diferencias remarcables entre las variantes MI(D)-BS 1 y MI(D)-BS 2.

Este patrón se observa tanto ante valores perdidos de mecanismo MCAR como ante valores perdidos de mecanismo MAR, siendo más acentuado en este último caso. Las diferencias mencionadas entre las estrategias BS-MI(D) y las estrategias MI(D)-BS se observan tanto si la imputación múltiple se lleva a cabo mediante el enfoque MI como si lleva a cabo mediante el enfoque MID; sin embargo, la tendencia a la sobrestimación de las estrategias MI(D)-BS es ligeramente más marcada bajo el primero de ellos.

En escenarios donde el EPV es especialmente reducido, si bien las estrategias BS-MI(D) estiman de forma más adecuada el valor del estadístico C, ninguna de las estrategias estudiadas corrige suficientemente el exceso de optimismo de las estimaciones aparentes; tampoco lo hace la validación bootstrap en ausencia de valores perdidos. Estos escenarios son aquellos en los que los modelos desarrollados presentan un mayor sobreajuste y una menor discriminación al ser empleados para la predicción en observaciones independientes.

Por el contrario, en escenarios con condiciones óptimas para el desarrollo de los modelos pronósticos, en los que se dispone de un EPV suficiente, las diferencias entre estrategias se

diluyen y el estadístico C se puede estimar adecuadamente mediante cualquiera de las estrategias estudiadas (Tabla 2). La variabilidad de las estimaciones se ve, asimismo, reducida.



**Figura 6.** Diferencia  $\Delta$  estadístico C: gráficos de caja por escenario y estrategia bajo valores perdidos de mecanismo MAR

**Tabla 2.** Diferencia  $\Delta$  estadístico C: valores medios por escenario y estrategia bajo valores perdidos de mecanismo MAR. Se utilizan tonos rosados para valores  $\bar{\Delta} > 0$ , correspondientes a una sobrestimación media del rendimiento; y tonos azulados para valores  $\bar{\Delta} < 0$ , correspondientes a una infraestimación media del rendimiento. La intensidad del color indica la distancia de estos valores respecto al cero.

Valores medios diferencia  $\Delta$  estadístico C (MAR)

n	propY	propmiss	EPV	MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID
100	0.10	0.02		0.0308	0.0306	0.0119	0.0263	0.0259	0.0123
		0.05	2	0.0503	0.0502	0.0072	0.0412	0.0412	0.0094
		0.10		0.0816	0.0819	0.0021	0.0673	0.0672	0.0052
	0.25	0.02		0.0129	0.0128	0.0053	0.0093	0.0094	0.0049
		0.05	5	0.0210	0.0206	0.0012	0.0142	0.0142	0.0024
		0.10		0.0415	0.0417	0.0022	0.0309	0.0306	0.0035
	0.50	0.02		0.0052	0.0053	0.0014	0.0032	0.0031	0.0015
		0.05	10	0.0093	0.0091	-0.0013	0.0039	0.0041	-0.0012
		0.10		0.0160	0.0159	-0.0061	0.0069	0.0067	-0.0054
500	0.10	0.02		0.0035	0.0036	-0.0016	0.0011	0.0010	-0.0025
		0.05	10	0.0082	0.0084	-0.0065	0.0034	0.0034	-0.0075
		0.10		0.0235	0.0233	-0.0080	0.0157	0.0156	-0.0085
	0.25	0.02		0.0009	0.0010	-0.0011	-0.0009	-0.0009	-0.0022
		0.05	25	0.0018	0.0018	-0.0030	-0.0013	-0.0013	-0.0044
		0.10		0.0057	0.0058	-0.0051	0.0006	0.0005	-0.0068
0.50	0.02		0.0007	0.0010	0.0000	-0.0003	-0.0004	-0.0009	
	0.05	50	0.0020	0.0019	-0.0001	-0.0009	-0.0008	-0.0015	
	0.10		0.0028	0.0030	-0.0025	-0.0013	-0.0014	-0.0042	
1000	0.10	0.02		0.0029	0.0030	-0.0001	0.0012	0.0010	-0.0011
		0.05	20	0.0069	0.0070	-0.0007	0.0030	0.0032	-0.0025
		0.10		0.0125	0.0124	-0.0054	0.0071	0.0070	-0.0063
	0.25	0.02		0.0007	0.0008	-0.0001	-0.0007	-0.0007	-0.0011
		0.05	50	0.0011	0.0010	-0.0014	-0.0017	-0.0016	-0.0031
		0.10		0.0037	0.0038	-0.0022	-0.0004	-0.0002	-0.0042
0.50	0.02		0.0000	0.0001	-0.0004	-0.0010	-0.0010	-0.0012	
	0.05	100	0.0001	0.0000	0.0012	0.0023	-0.0024	-0.0029	
	0.10		0.0014	0.0014	-0.0014	-0.0019	-0.0021	-0.0037	

### 7.1.3.2. Calibración

#### 7.1.3.2.1. *Calibration slope*

Las tendencias observadas en relación con la estimación del estadístico C se presentan de forma similar en los resultados correspondientes a la pendiente de calibración o *calibration slope* (Figura 7).

Las estrategias BS-MI(D) realizan una mayor corrección del optimismo presente en las estimaciones aparentes de la pendiente de calibración. En algunos escenarios, estas estrategias presentan una ligera tendencia a la infraestimación; no obstante, en la práctica totalidad de los escenarios estudiados, proporcionan estimaciones más próximas en media a los verdaderos valores de la pendiente de calibración que las estrategias MI(D)-BS (Tabla 3). Las estrategias MI(D)-BS, por el contrario, presentan una mayor tendencia a la sobrestimación, especialmente en escenarios de menor EPV. Las variantes MI(D)-BS 1 y MI(D)-BS 2 vuelven a presentar resultados similares.

Las diferencias entre estrategias se encuentran tanto ante valores perdidos de mecanismo MCAR como ante valores perdidos de mecanismo MAR, si bien en este último caso se hacen más notorias. Se presentan bajo los dos enfoques empleados para la imputación múltiple, MI y MID, siendo ligeramente más marcadas bajo el enfoque MI.

Nuevamente, se observa la importancia del número de eventos por variable para la adecuada estimación del rendimiento mediante la validación interna. En escenarios con EPV especialmente reducido, todas las estrategias evaluadas tienden a sobrestimar la pendiente de calibración; incluso la validación bootstrap en ausencia de valores perdidos. En decir, la pendiente de calibración se estima de forma menos fiable justamente en aquellos escenarios en los que modelos pronósticos resultan pobremente calibrados y requieren recalibración.

MAR

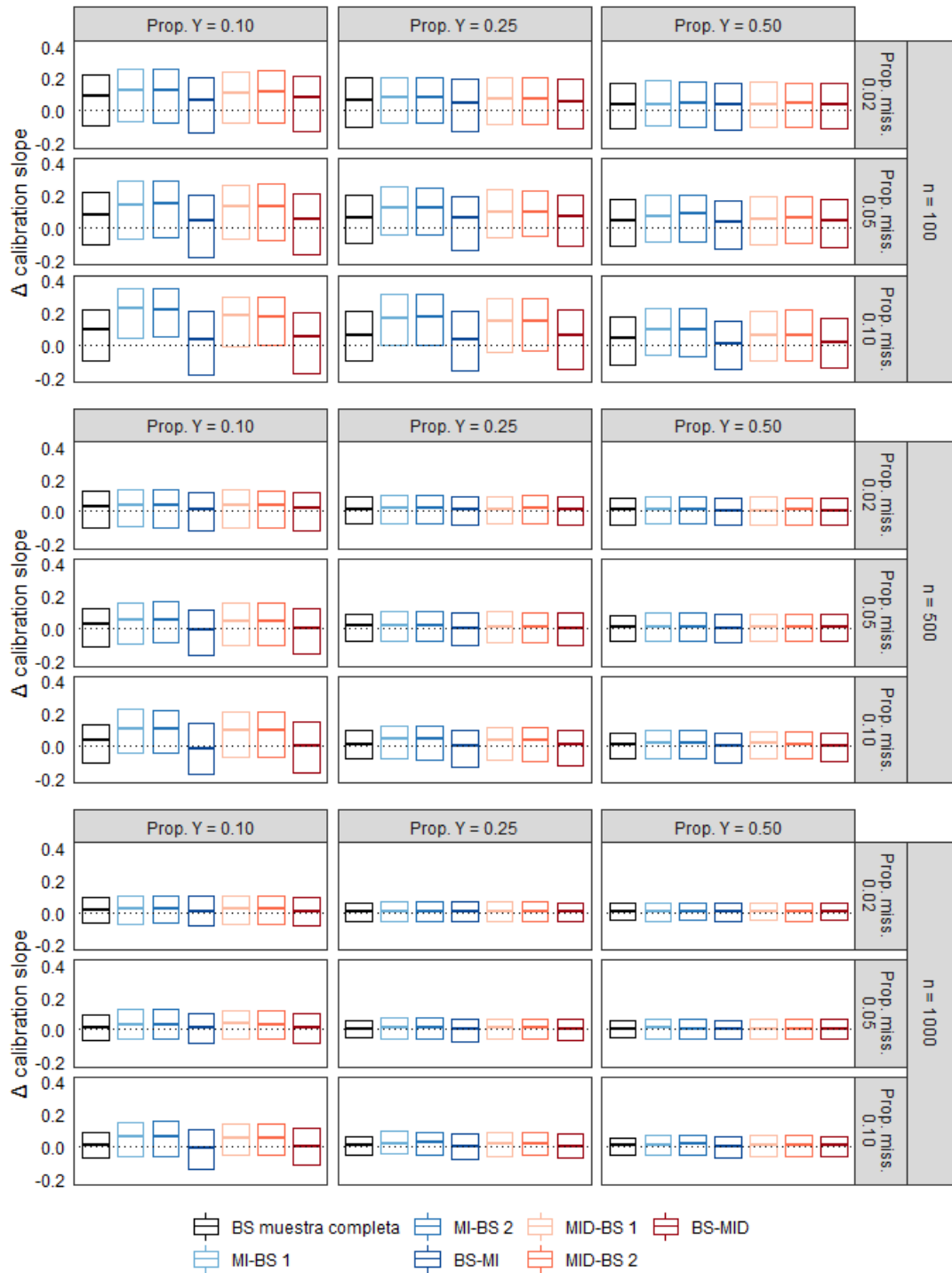


Figura 7. Diferencia  $\Delta$  calibration slope: gráficos de caja por escenario y estrategia bajo valores perdidos de mecanismo MAR

**Tabla 3.** Diferencia  $\Delta calibration slope$ : valores medios por escenario y estrategia bajo valores perdidos de mecanismo MAR. Se utilizan tonos rosados para valores  $\bar{\Delta} > 0$ , correspondientes a una sobrestimación media del rendimiento; y tonos azulados para valores  $\bar{\Delta} < 0$ , correspondientes a una infraestimación media del rendimiento. La intensidad del color indica la distancia de estos valores respecto al cero.

Valores medios diferencia  $\Delta calibration slope$  (MAR)

n	propY	propmiss	EPV	MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID
100	0.10	0.02	2	0.0501	0.0501	-0.0026	0.0415	0.0402	0.0050
		0.05		0.0943	0.0930	-0.0148	0.0693	0.0691	-0.0063
		0.10		0.1673	0.1670	-0.0270	0.1267	0.1254	-0.0077
	0.25	0.02	5	0.0393	0.0395	0.0125	0.0300	0.0305	0.0156
		0.05		0.0595	0.0576	-0.0110	0.0408	0.0409	0.0011
		0.10		0.1249	0.1259	-0.0072	0.0957	0.0953	0.0083
	0.50	0.02	10	0.0093	0.0104	-0.0064	0.0071	0.0065	-0.0007
		0.05		0.0246	0.0234	-0.0170	0.0110	0.0122	-0.0081
		0.10		0.0499	0.0493	-0.0350	0.0251	0.0240	-0.0220
500	0.10	0.02	10	0.0054	0.0055	-0.0152	0.0006	0.0005	-0.0130
		0.05		0.0127	0.0130	-0.0441	0.0032	0.0029	-0.0377
		0.10		0.0675	0.0671	-0.0500	0.0536	0.0529	-0.0340
	0.25	0.02	25	0.0040	0.0046	-0.0040	0.0024	0.0021	-0.0030
		0.05		0.0053	0.0056	-0.0149	0.0011	0.0010	-0.0118
		0.10		0.0199	0.0202	-0.0242	0.0101	0.0097	-0.0191
0.50	0.02	50	0.0008	0.0020	-0.0021	0.0017	0.0012	-0.0008	
	0.05		0.0061	0.0052	-0.0035	0.0027	0.0033	0.0005	
	0.10		0.0093	0.0101	-0.0137	0.0030	0.0029	-0.0090	
1000	0.10	0.02	20	0.0106	0.0108	-0.0029	0.0096	0.0087	0.0001
		0.05		0.0276	0.0278	-0.0021	0.0225	0.0235	0.0021
		0.10		0.0424	0.0423	-0.0279	0.0352	0.0348	-0.0165
	0.25	0.02	50	0.0032	0.0035	0.0003	0.0018	0.0018	0.0006
		0.05		0.0059	0.0057	-0.0047	0.0024	0.0031	-0.0033
		0.10		0.0171	0.0173	-0.0071	0.0117	0.0126	-0.0033
0.50	0.02	100	0.0019	0.0025	0.0001	0.0026	0.0027	0.0020	
	0.05		0.0023	0.0017	-0.0036	0.0005	0.0002	-0.0023	
	0.10		0.0081	0.0081	-0.0042	0.0055	0.0049	-0.0019	

#### 7.1.3.2.2. *Calibration-in-the-large (CITL)*

Los resultados relativos a la estimación de la CITL se presentan en la Figura 8 y en la Tabla 4. En los escenarios de tamaño muestral  $n = 100$  y proporción muestral de eventos  $\text{prop}Y = 0.10$ , se han encontrado frecuentes problemas de convergencia en la estimación de la CITL. Los resultados reportados corresponden, por lo tanto, a un número reducido de simulaciones, lo que los hace menos robustos y de difícil comparación (Anexo 17. Diferencia  $\Delta$  CITL por estrategias: estadísticos descriptivos bajo MCAR, Anexo 18. Diferencia  $\Delta$  CITL por estrategias: estadísticos descriptivos bajo MAR).

En los escenarios restantes, pueden observarse ciertas diferencias entre estrategias, si bien no se identifica una tendencia clara a favor de alguna de ellas.

A diferencia de lo que ocurre en el caso del estadístico C y la pendiente de calibración, donde valores estimados por encima del valor de referencia correspondiente ( $\Delta > 0$ ) sugieren una estimación excesivamente optimista, la interpretación de los resultados para la CITL resulta menos intuitiva. A pesar de esta limitación, se observa una elevada variabilidad en los valores  $\Delta$  cuando los modelos se desarrollan en circunstancias poco óptimas; que se reduce en escenarios de mayor EPV. Esta variabilidad, en términos de rango intercuartílico, es ligeramente superior bajo las estrategias BS-MI(D).

MAR

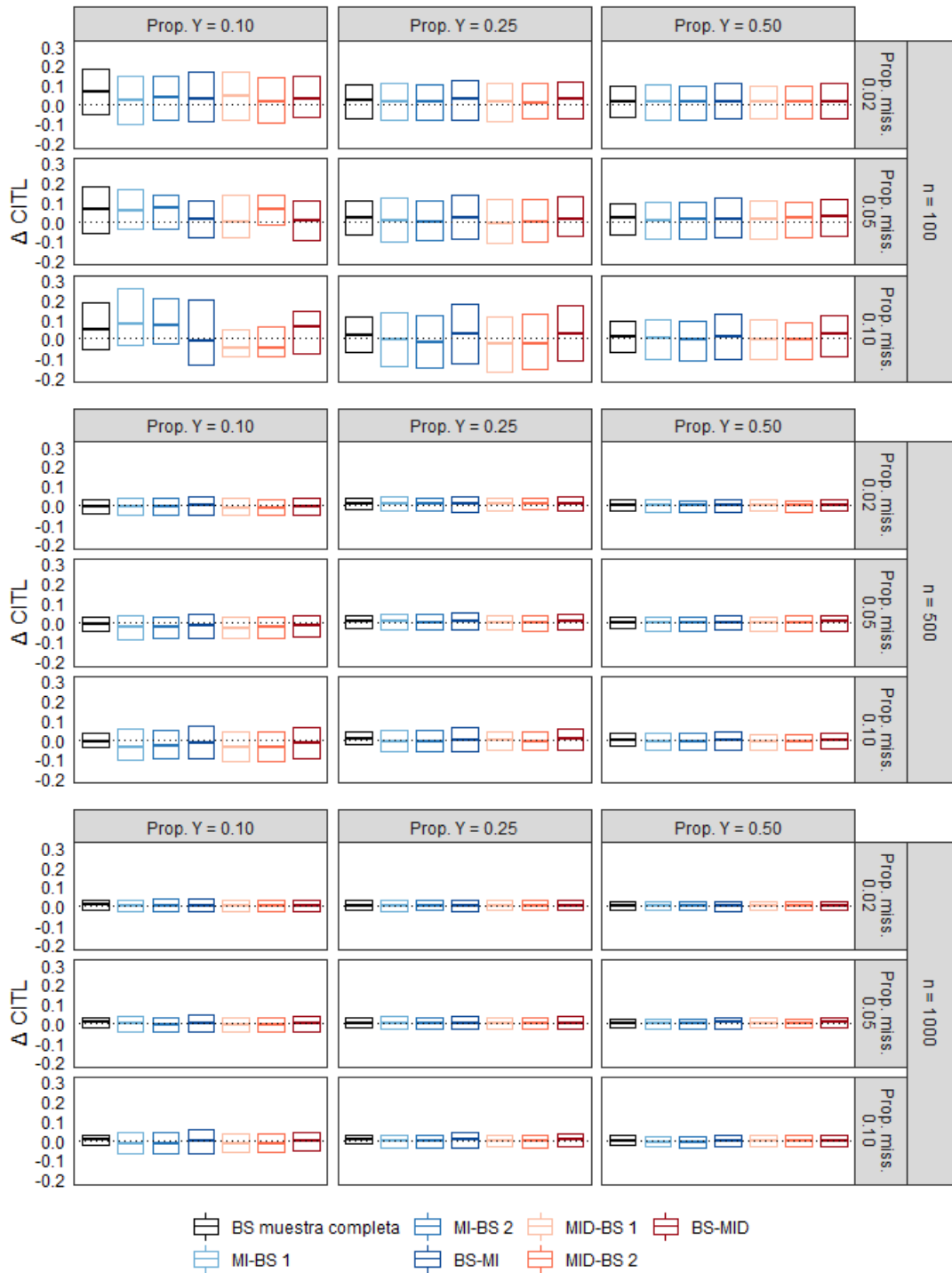


Figura 8. Diferencia  $\Delta$  CITL: gráficos de caja por escenario y estrategia bajo valores perdidos de mecanismo MAR

**Tabla 4.** Diferencia  $\Delta$  CITL: valores medios por escenario y estrategia bajo valores perdidos de mecanismo MAR. Se utilizan tonos rosados para valores  $\bar{\Delta} > 0$ , correspondientes a una sobrestimación media del rendimiento; y tonos azulados para valores  $\bar{\Delta} < 0$ , correspondientes a una infraestimación media del rendimiento. La intensidad del color indica la distancia de estos valores respecto al cero.

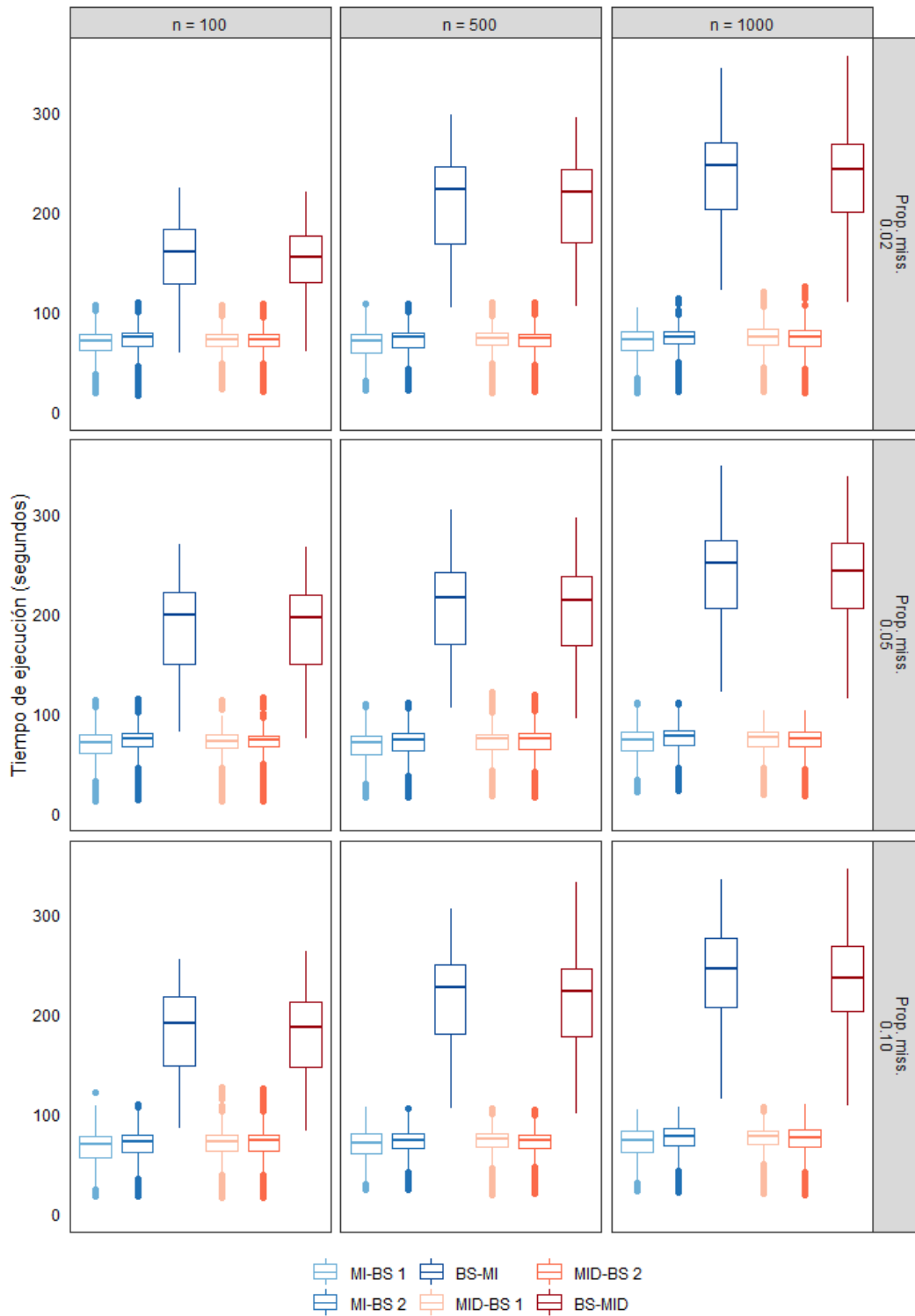
Valores medios diferencia $\Delta$ CITL (MAR)										
n	propY	propmiss	EPV	MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID	
100	0.10	0.02	2	0.0203	0.0520	0.0471	0.0335	0.0337	0.0423	
		0.05		0.0578	0.0481	0.0089	0.0211	0.0482	0.0148	
		0.10		0.1062	0.1059	0.0321	0.0057	-0.0223	0.0039	
	0.25	0.02	5	0.0137	0.0126	0.0186	0.0111	0.0128	0.0212	
		0.05		0.0138	0.0099	0.0270	0.0041	0.0080	0.0314	
		0.10		-0.0038	-0.0054	0.0361	-0.0184	-0.0124	0.0340	
	0.50	0.02	10	0.0106	0.0111	0.0134	0.0139	0.0123	0.0153	
		0.05		0.0062	0.0074	0.0191	0.0086	0.0103	0.0224	
		0.10		-0.0082	-0.0095	0.0136	-0.0046	-0.0070	0.0185	
	500	0.10	0.02	10	-0.0095	-0.0094	-0.0051	-0.0104	-0.0111	-0.0070
			0.05		-0.0266	-0.0276	-0.0200	-0.0284	-0.0286	-0.0185
			0.10		-0.0338	-0.0341	-0.0196	-0.0393	-0.0376	-0.0162
0.25		0.02	25	0.0045	0.0046	0.0052	0.0046	0.0043	0.0053	
		0.05		-0.0012	-0.0021	0.0019	-0.0033	-0.0028	0.0019	
		0.10		-0.0088	-0.0095	-0.0004	-0.0114	-0.0117	-0.0020	
0.50	0.02	50	-0.0032	-0.0038	-0.0033	-0.0031	-0.0031	-0.0022		
	0.05		-0.0062	-0.0057	-0.0053	-0.0057	-0.0044	-0.0033		
	0.10		-0.0108	-0.0103	-0.0060	-0.0091	-0.0100	-0.0043		
1000	0.10	0.02	20	0.0004	0.0008	0.0023	0.0016	0.0009	0.0033	
		0.05		-0.0072	-0.0083	-0.0033	-0.0094	-0.0093	-0.0039	
		0.10		-0.0123	-0.0123	-0.0034	-0.0151	-0.0147	-0.0026	
	0.25	0.02	50	0.0012	0.0014	0.0026	0.0014	0.0015	0.0022	
		0.05		-0.0001	-0.0003	0.0015	-0.0004	-0.0012	0.0013	
		0.10		-0.0031	-0.0028	0.0014	-0.0033	-0.0031	0.0022	
0.50	0.02	100	-0.0004	-0.0001	-0.0006	0.0003	0.0002	0.0007		
	0.05		-0.0020	-0.0016	0.0000	-0.0014	-0.0016	-0.0005		
	0.10		-0.0048	-0.0050	-0.0027	-0.0040	-0.0039	-0.0010		

## 7.2. Tiempo computacional

Los resultados del tiempo de ejecución requerido para la aplicación de las distintas estrategias evaluadas en el estudio de simulación se presentan en función del tamaño muestral y de la proporción de valores perdidos por variable.

En comparación con las estrategias MI(D)-BS, las estrategias BS-MI(D) requieren un tiempo de ejecución claramente superior (Figura 9). Este tiempo de ejecución aumenta conforme aumenta el tamaño muestral. En las muestras de menor tamaño, se observa también un incremento temporal cuando la cantidad de información perdida es mayor; en muestras de mayor tamaño, el impacto de la proporción de valores perdidos por variable sobre el tiempo de ejecución deja de notarse. Las estrategias MI(D)-BS presentan tiempos de ejecución no sólo considerablemente menores, sino también más robustos ante variaciones en el tamaño muestral y en la proporción de valores perdidos por variable.

Los tiempos medios de ejecución por escenario y estrategia se presentan en la Tabla 5. En ella se recogen también los tiempos medios relativos, obtenidos mediante cociente sobre el tiempo mínimo de ejecución por escenario. En relación con las estrategias MI(D)-BS, las estrategias BS-MI(D) requieren tiempos de ejecución entre 2 y 3.5 veces mayores en media. Las variantes MI(D)-BS 1 y MI(D)-BS 2 requieren tiempos de ejecución similares.



**Figura 9.** Tiempos de ejecución por estrategia (segundos) en función del tamaño muestral y de la proporción de valores perdidos por variable

**Tabla 5.** Tiempos medios de ejecución por estrategia, en segundos y relativos. Los tiempos medios relativos se obtienen como cociente sobre el tiempo mínimo bajo el escenario correspondiente; los resultados se obtienen, de forma separada, para los procedimientos de imputación MI y MID.

		Tiempos medios de ejecución por estrategia						
n	propmiss		MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID
100	0.02	Segundos	68.4072	70.9934	155.5030	69.6570	68.8993	152.6084
		Relativo	Mín.	1.04	2.27	1.01	Mín.	2.21
	0.05	Segundos	67.5178	71.0499	187.8972	69.1232	69.1428	184.4962
		Relativo	Mín.	1.05	2.78	1.00	Mín.	2.67
	0.10	Segundos	65.7230	69.1126	183.9207	69.6696	69.8036	181.7229
		Relativo	Mín.	1.05	2.80	1.00	Mín.	2.61
500	0.02	Segundos	68.3757	70.5682	210.9639	70.1825	69.9649	208.4085
		Relativo	Mín.	1.03	3.09	1.00	Mín.	2.98
	0.05	Segundos	67.3894	70.5860	208.1446	70.2479	70.2063	205.5555
		Relativo	Mín.	1.05	3.09	1.00	Mín.	2.93
	0.10	Segundos	69.4756	71.1002	216.7372	70.1919	69.4697	212.9280
		Relativo	Mín.	1.02	3.12	1.01	Mín.	3.07
1000	0.02	Segundos	69.7182	71.6814	240.3656	72.4858	71.9286	237.4845
		Relativo	Mín.	1.03	3.45	1.01	Mín.	3.30
	0.05	Segundos	70.4564	73.3480	240.3299	71.8491	71.0591	236.3168
		Relativo	Mín.	1.04	3.41	1.01	Mín.	3.33
	0.10	Segundos	71.2044	75.1676	240.2658	73.4350	73.0916	235.1200
		Relativo	Mín.	1.06	3.37	1.00	Mín.	3.22

En las estrategias MI(D)-BS, la imputación de los valores perdidos se realiza **M** veces sobre la muestra inicial y, posteriormente, se seleccionan **B** muestras bootstrap a partir de cada muestra imputada (es decir, se seleccionan **B · M** muestras bootstrap en total). En las estrategias BS-MI(D), se seleccionan inicialmente **B** muestras bootstrap y sobre cada una de ellas se lleva a cabo la imputación de los valores perdidos **M** veces; se trata de estrategias en las que se realiza una aplicación más intensiva del procedimiento de imputación múltiple de valores perdidos, implementado **B · M** veces en total, lo que explica su mayor tiempo de computación.

Debe notarse que los tiempos de ejecución requeridos por cada una de las estrategias evaluadas dependen del número de muestras bootstrap seleccionadas (**B**) y del número de imputaciones realizadas (**M**), así como del número de iteraciones llevadas a cabo al implementar el algoritmo MICE. Por otra parte, la complejidad que adquiera el análisis estadístico de interés, así como la elección de los métodos de imputación univariantes a integrar en el algoritmo MICE, pueden derivar en un incremento considerable de los tiempos de computación.

## 8. Discusión

En el presente trabajo se han evaluado, mediante simulación, diferentes estrategias para la integración de la imputación múltiple de valores perdidos en la validación bootstrap de modelos pronósticos. Las estrategias se han evaluado en el marco de la regresión logística binomial, considerando valores perdidos tanto en los predictores  $X$  como en la variable de interés  $Y$ . La comparación de las estrategias se ha realizado bajo dos posibles enfoques de imputación múltiple cuando la variable de interés se presenta incompleta: los enfoques MI y MID.

En relación con el estadístico  $C$  y la pendiente de calibración, los resultados de este trabajo muestran que las estrategias BS-MI(D) proporcionan estimaciones de rendimiento más realistas; mientras que las estrategias MI(D)-BS tienden a sobrestimar el rendimiento predictivo. Las diferencias entre estrategias son más notorias ante valores perdidos de mecanismo MAR y se encuentran, fundamentalmente, en escenarios de reducido EPV. Cuando los modelos pronósticos se desarrollan bajo un EPV suficiente, las diferencias entre estrategias se diluyen.

Las tendencias observadas son similares bajo los enfoques MI y MID de imputación múltiple en presencia de valores perdidos en  $Y$ . No obstante, la tendencia a la sobrestimación de las estrategias MI(D)-BS es ligeramente más pronunciada si se emplea el enfoque MI. Bajo este enfoque, observaciones en las que el valor de la variable  $Y$  es imputado a partir de la información proporcionada por los predictores  $X$  se emplean no sólo en el desarrollo del modelo pronóstico, sino también en su validación interna, lo que puede redundar en un mayor optimismo en la estimación del rendimiento.

En relación con la CITL, los resultados obtenidos no resultan concluyentes.

La cuestión relativa a cómo debe integrarse la imputación múltiple de valores perdidos en la validación bootstrap de modelos pronósticos ha sido estudiada previamente mediante simulación por Musoro et al. (6) y por Wahl et al. (7). Los resultados de estas investigaciones indican un menor sesgo en la estimación del rendimiento predictivo cuando la imputación

múltiple de valores perdidos se lleva a cabo tras el remuestreo bootstrap; coincidiendo, en este aspecto, con los resultados del presente trabajo.

Los estudios de Musoro et al. (6) y Wahl et al. (7) consideran valores perdidos únicamente en los predictores de los modelos pronósticos. En este sentido, el presente trabajo estudia la problemática de interés en una situación no abordada previamente, caracterizada por que la variable de interés en la modelización también se presenta incompleta.

Los problemas de convergencia encontrados en la estimación de la CITL suponen una limitación de este trabajo; son necesarios futuros estudios para evaluar en mayor profundidad el comportamiento de las diferentes estrategias para la estimación de esta medida de calibración.

Cuestiones de interés a ser abordadas en futuros trabajos son la comparación de estrategias cuando se implementan métodos de selección de predictores en el desarrollo de los modelos pronósticos; así como el estudio de variantes existentes de validación bootstrap, *out-of-bag* y  $0.632+$ .

## 9. Conclusiones

En la validación bootstrap de modelos pronósticos desarrollados en presencia de valores perdidos que afectan tanto a los predictores como a la variable de interés, las estrategias BS-MI(D) resultan preferibles para la integración de las técnicas de imputación múltiple. Estas estrategias, no obstante, requieren un mayor tiempo de ejecución. Las estrategias MI(D)-BS resultan adecuadas cuando los modelos pronósticos se desarrollan bajo un EPV suficientemente elevado, y pueden considerarse en estos escenarios si la capacidad computacional supone una limitación.

## 10. Bibliografía

1. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Biol* [Internet]. 2013;11:EE. Available from: [www.annals.org](http://www.annals.org)
2. Steyerberg EW. *Clinical Prediction Models. A practical approach to development, validation and updating.* New York: Springer; 2009.
3. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001;54(8):774–81.
4. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol.* 2003;56(5):441–7.
5. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3):1–67.
6. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. *BMC Med Res Methodol* [Internet]. 2014 [cited 2022 Apr 27];14:116. Available from: <http://www.biomedcentral.com/1471-2288/14/116>
7. Wahl S, Boulesteix AL, Zierer A, Thorand B, Avan de Wiel M. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. *BMC Med Res Methodol.* 2016 Oct 26;16(1):1–18.
8. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res.* 2017;26(2):796–808.
9. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* [Internet]. 1984 Apr;3(2):143–52. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/sim.4780030207>

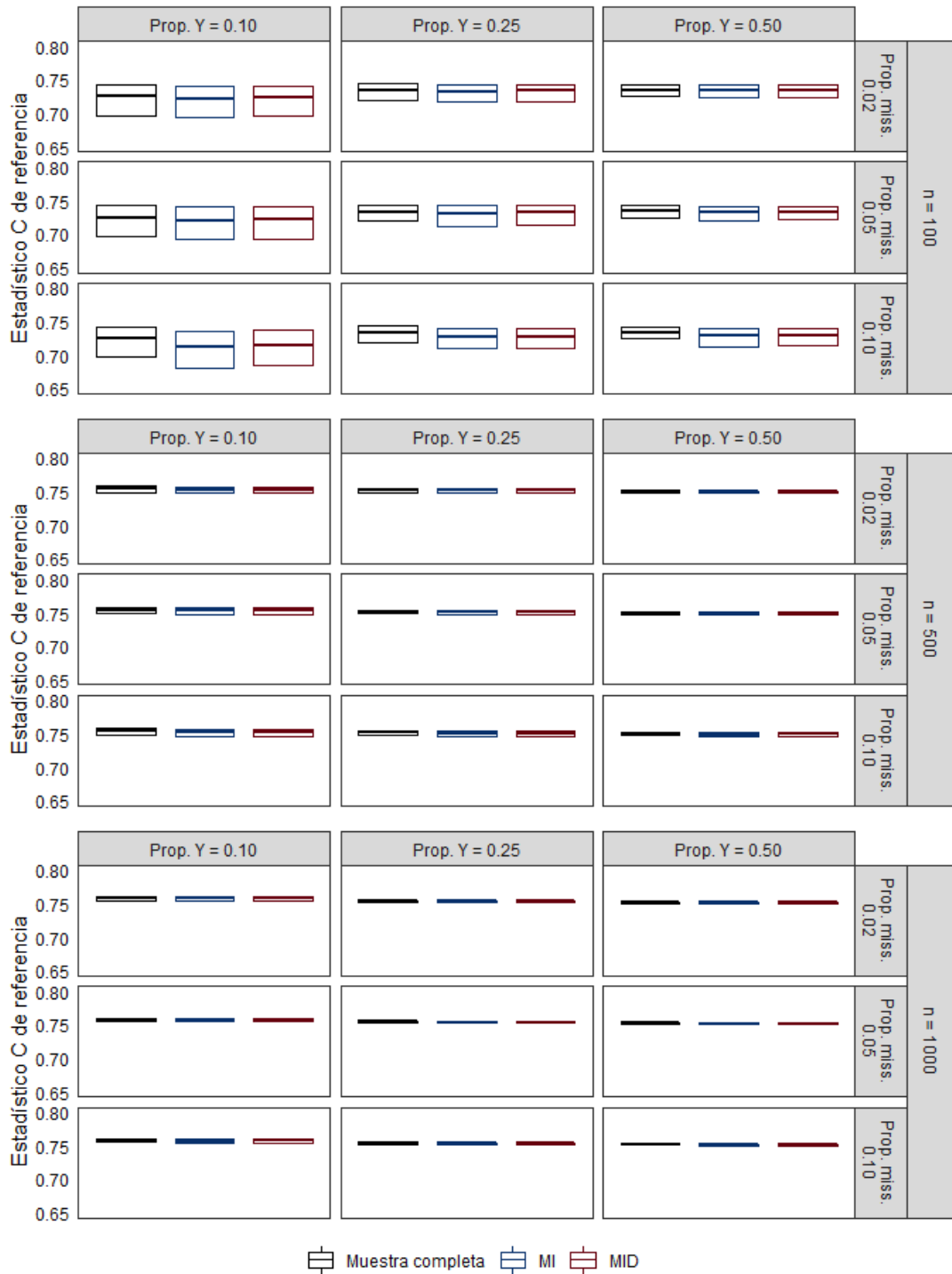
10. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* [Internet]. 1996 Dec;49(12):1373–9. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0895435696002363>
11. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Chapman and Hall; 1993.
12. Efron B. Estimating the error rate of a prediction rule: some improvements on cross-validation. *J Am Stat Assoc*. 1983;(78):316–31.
13. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–92.
14. Lee KJ, Tilling KM, Cornish RP, Little RJA, Bell ML, Goetghebeur E, et al. Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework. *J Clin Epidemiol*. 2021;134:79–88.
15. Hughes RA, Heron J, Sterne JAC, Tilling KM. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol*. 2019;48(4):1294–304.
16. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Ltd; 1987.
17. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci* [Internet]. 2007 Sep [cited 2022 May 6];8(3):206–13. Available from: <https://pubmed.ncbi.nlm.nih.gov/17549635/>
18. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res* [Internet]. 2007 Jun [cited 2022 May 6];16(3):199–218. Available from: <https://pubmed.ncbi.nlm.nih.gov/17621468/>
19. Vink G, Frank LE, Pannekoek J, Van Buuren S. Predictive mean matching imputation of semicontinuous variables. 2014;
20. Van Buuren S, L Brand JP, M Groothuis-oudshoorn CG, Rubin DB. Fully conditional specification in multivariate imputation. *J Stat Comput Simul* [Internet]. 2006 [cited 2022 May 6];76(12):1049–64. Available from: <http://www.tandf.co.uk/journals>

21. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006 Oct;59(10):1092–101.
22. Von Hippel PT. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociol Methodol*. 2007 Dec;37(1):83–117.
23. Harrell FE, Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Tutorials Biostat Stat Methods Clin Stud*. 2005;1:223–49.
24. Harrell FE. *Regression Modeling Strategies*. New York: Springer; 2001.
25. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: The Achilles heel of predictive analytics. *BMC Med*. 2019 Dec 16;17(1).
26. Riley RD, van der Windt DA, Croft P, Moons KGM. *Prognosis research in healthcare*. Oxford; 2019.
27. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the “calibration slope” really measure? *J Clin Epidemiol* [Internet]. 2020 Feb 1 [cited 2022 Apr 12];118:93–9. Available from: <https://doi.org/10.1016/j.jclinepi.2019.09.016>
28. Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema DF. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med*. 2000;(19):1059–79.
29. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study [Internet]. 2010. Available from: <http://www.biomedcentral.com/1471-2288/10/7>

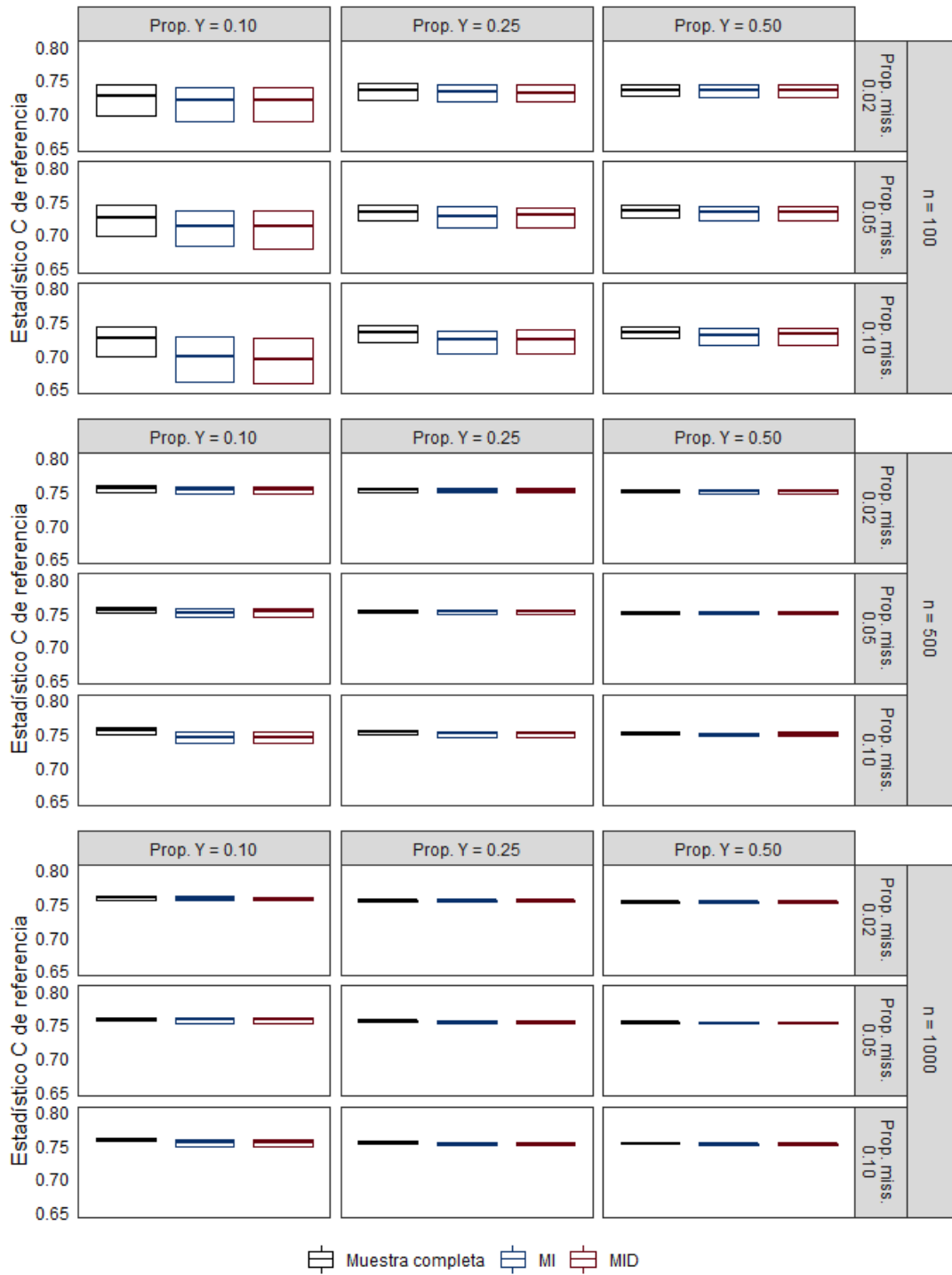
## 11. Anexos

### Anexo 1. Estadístico C de referencia

#### MCAR

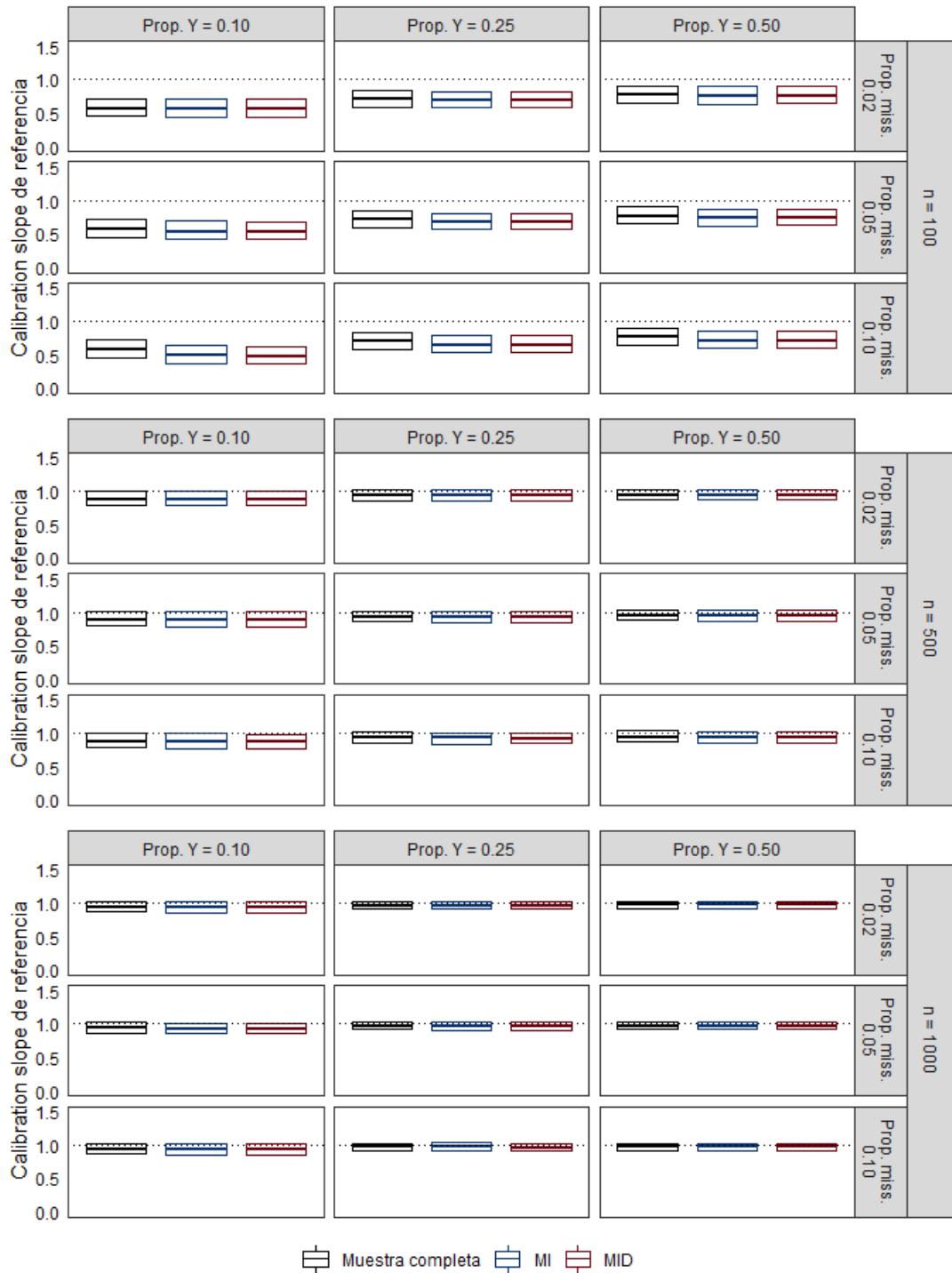


MAR

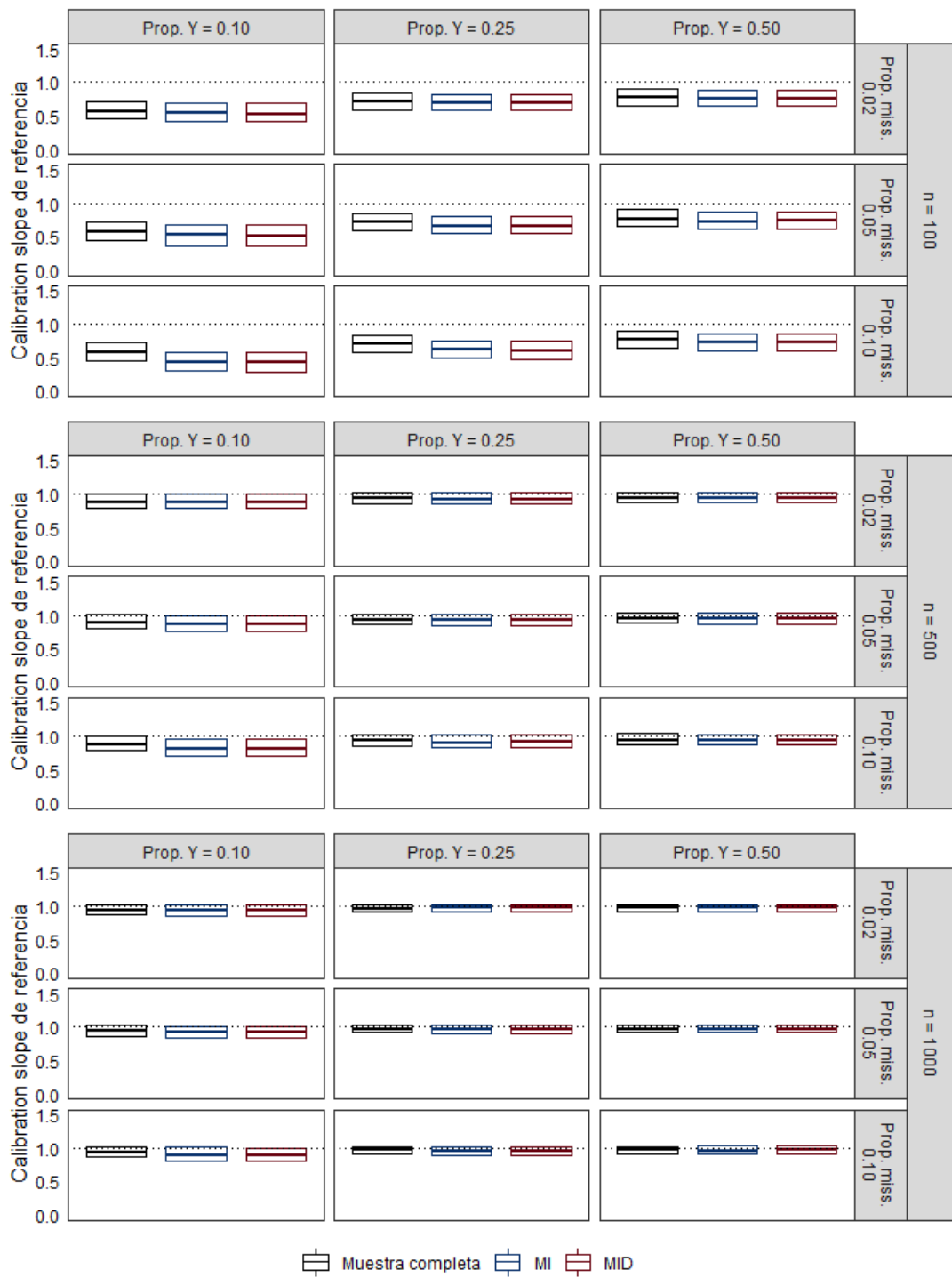


Anexo 2. Calibration slope de referencia

MCAR

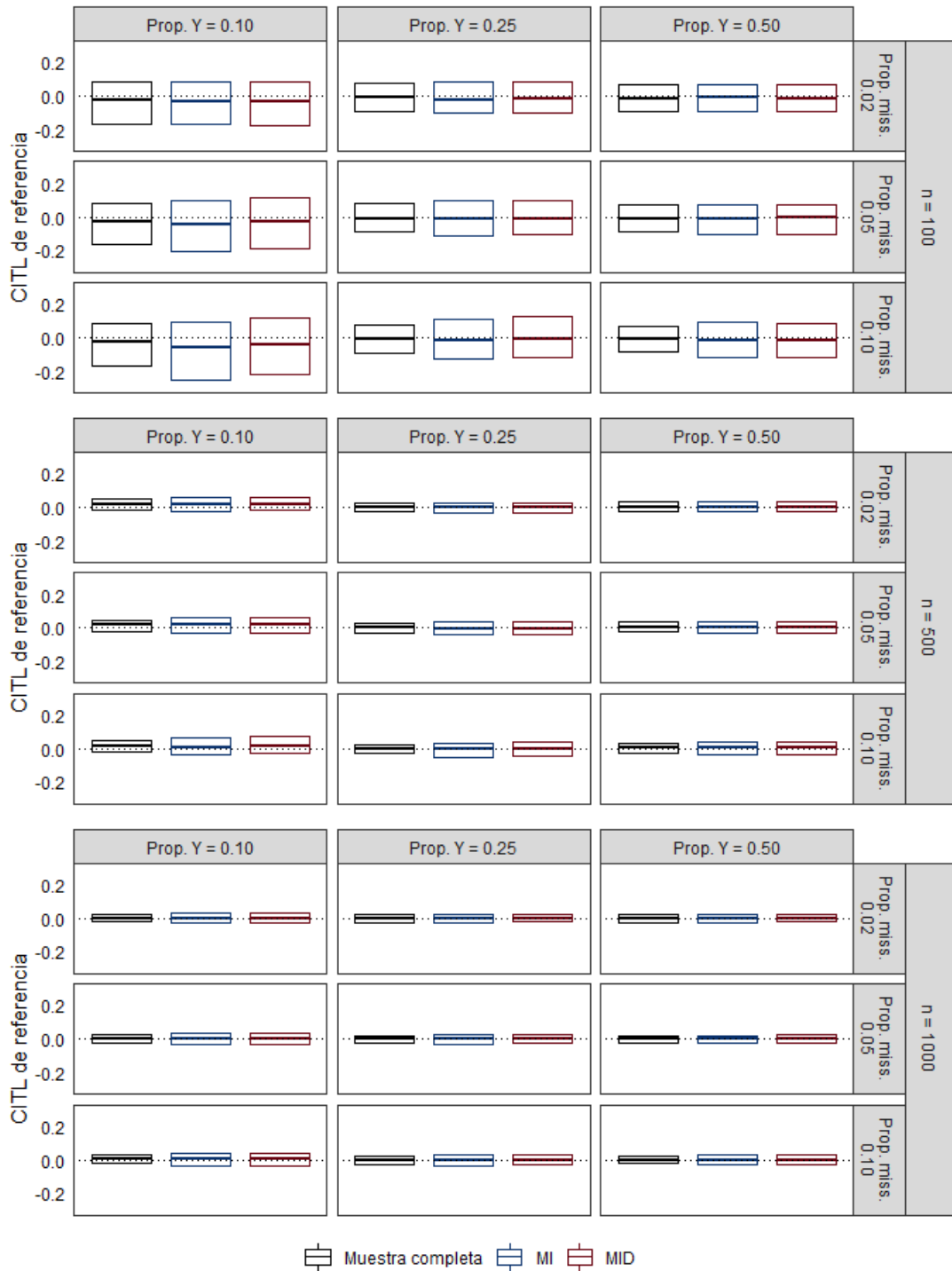


MAR

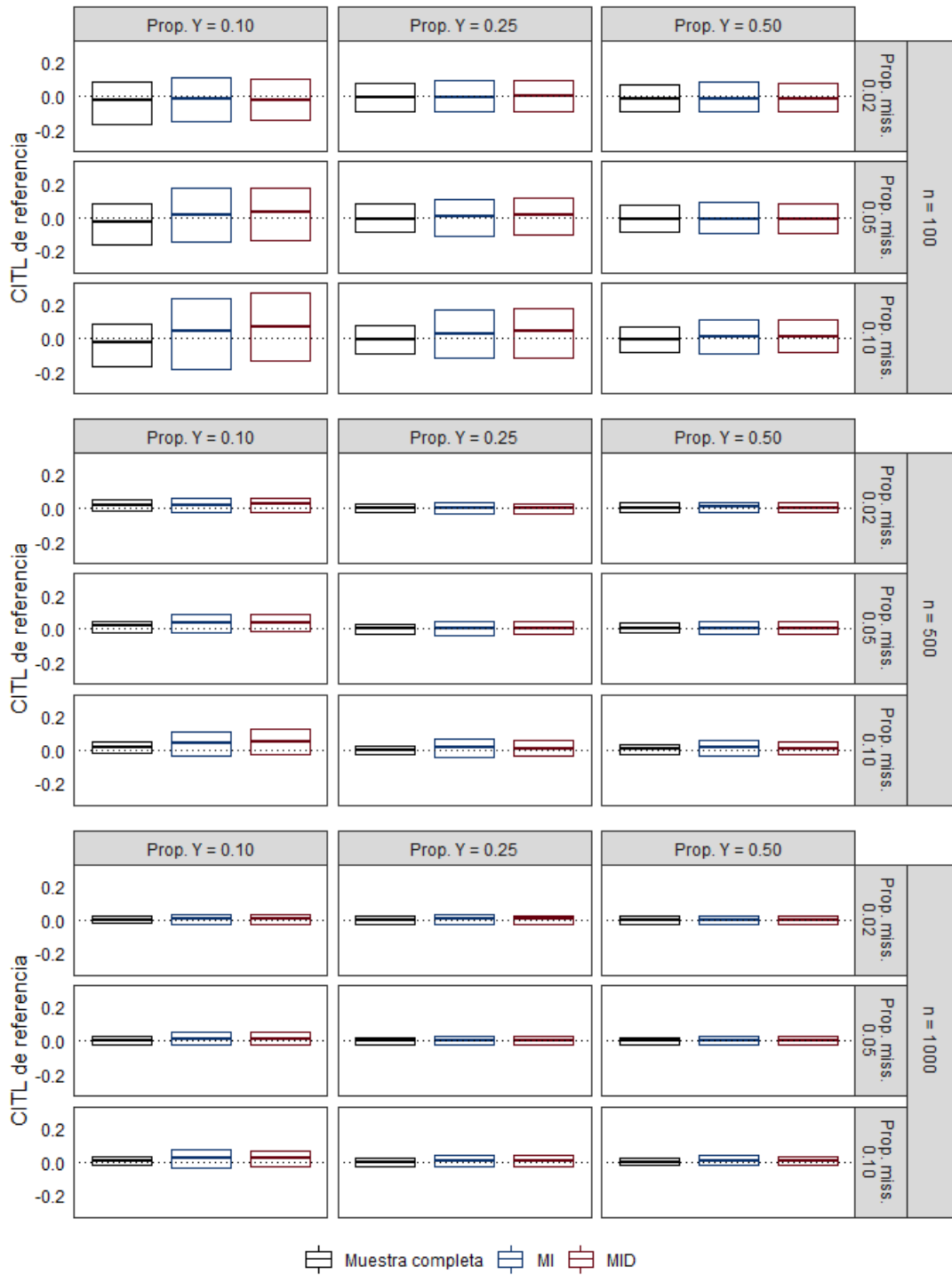


Anexo 3. Calibration-in-the-large de referencia

MCAR

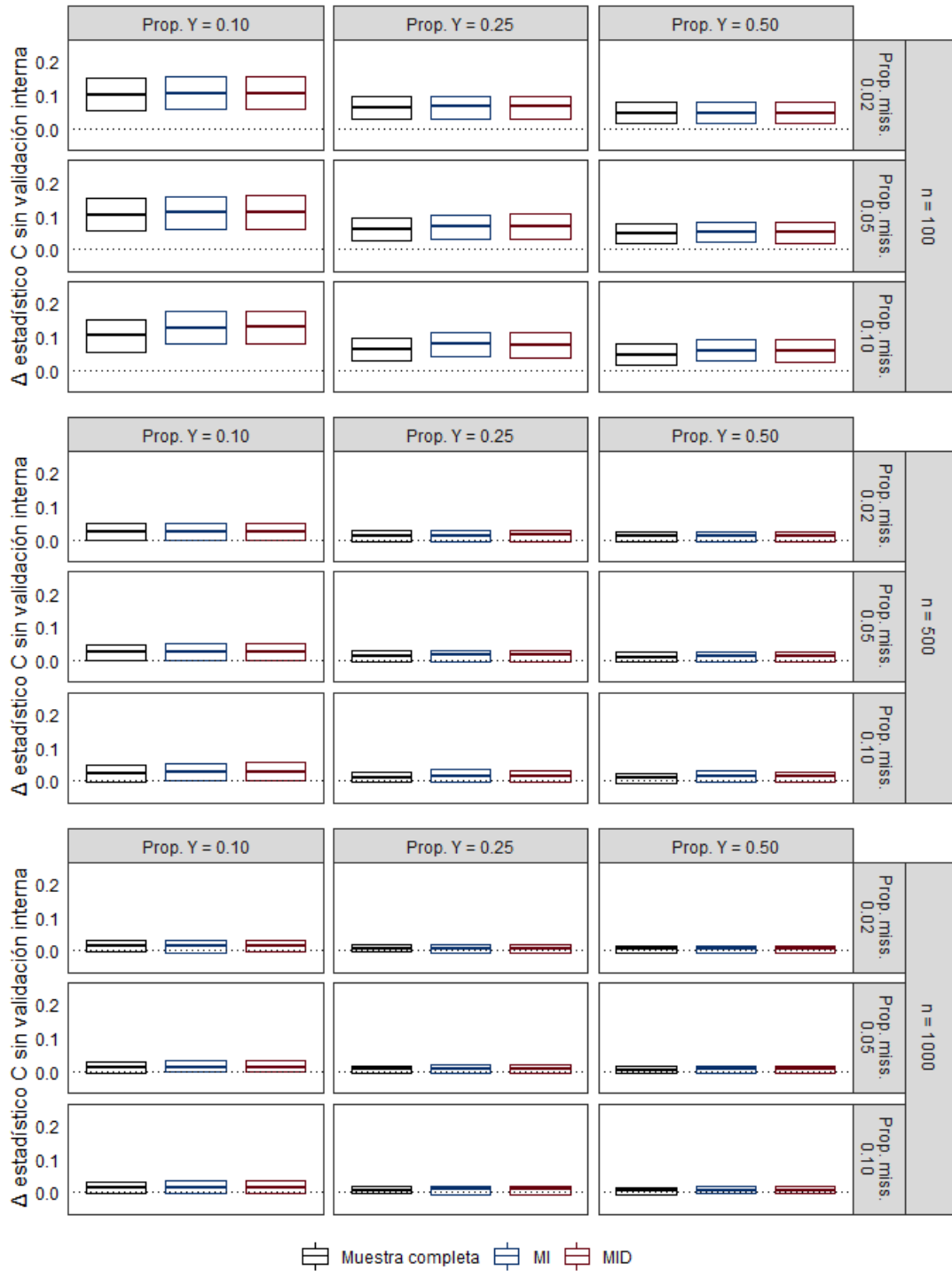


MAR

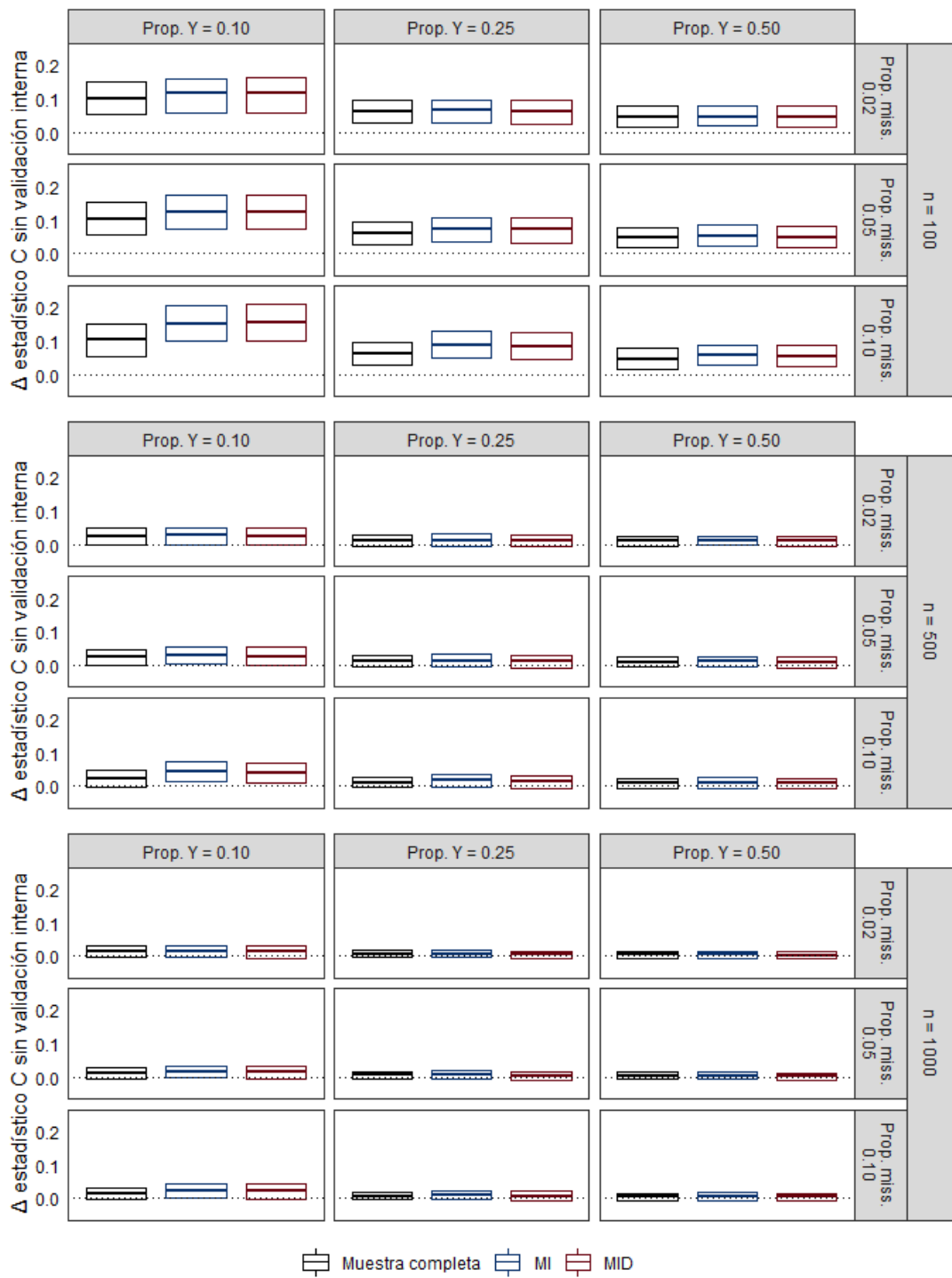


Anexo 4. Diferencia  $\Delta$  estadístico C sin validación interna

MCAR

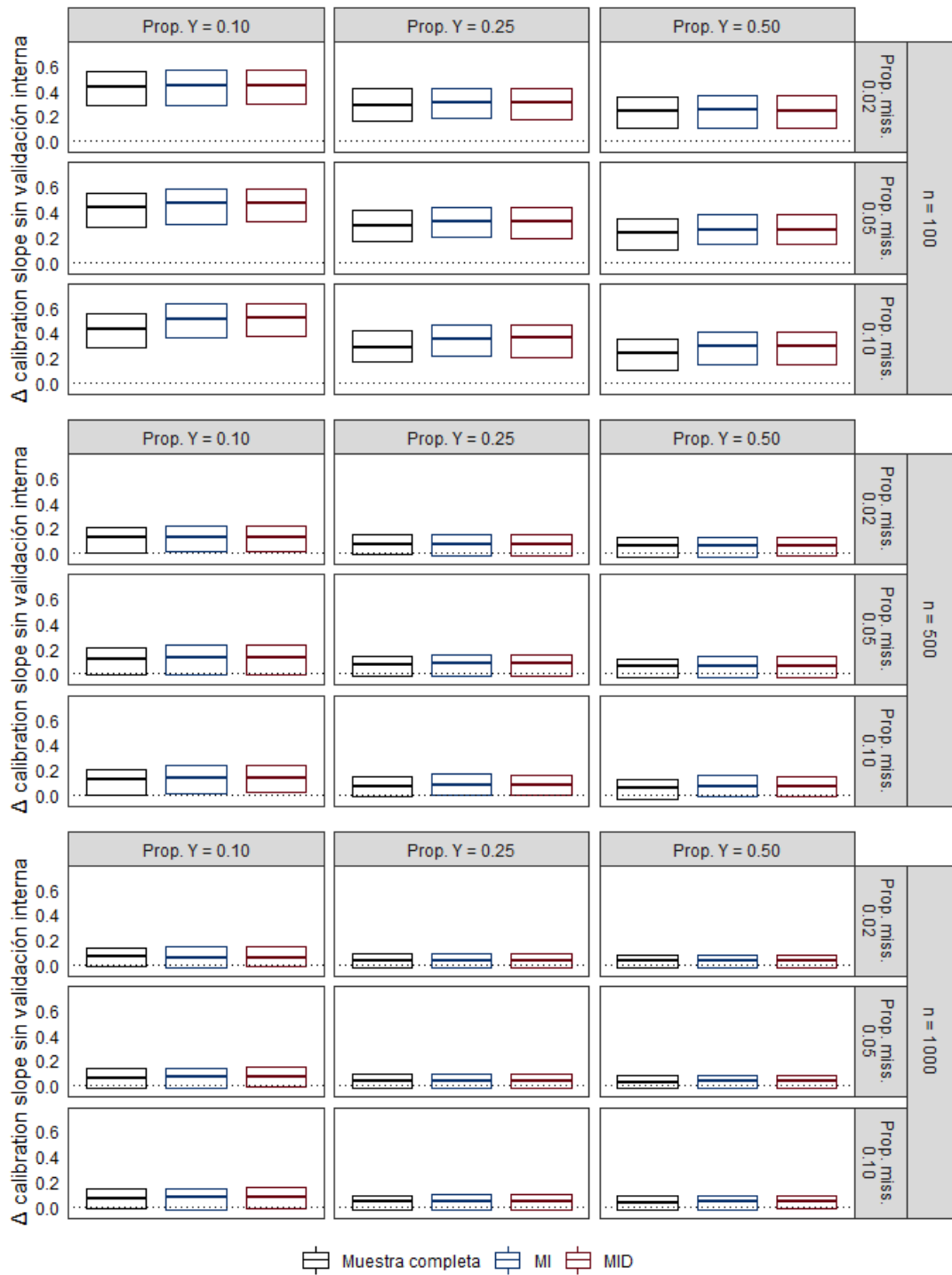


MAR

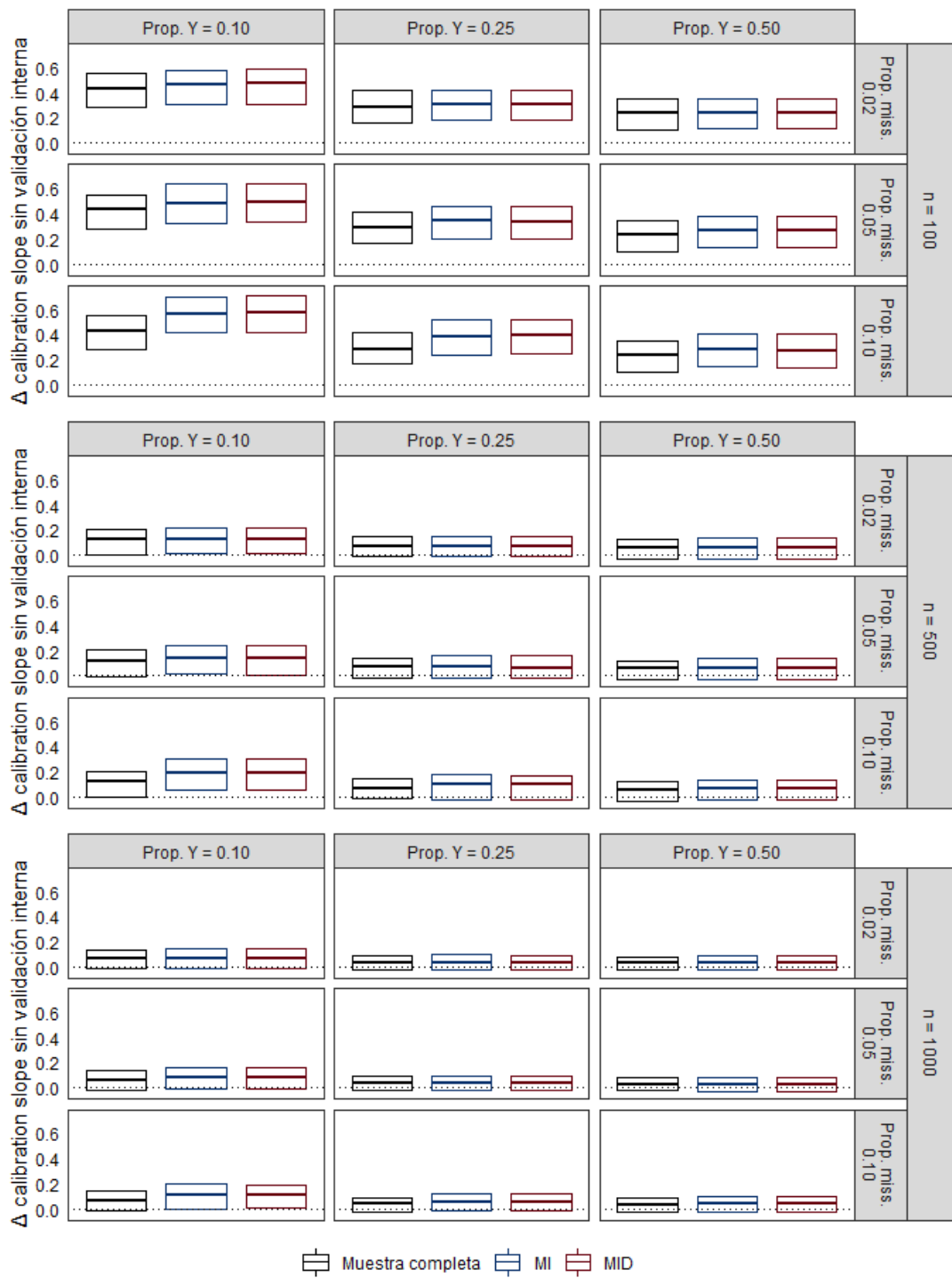


Anexo 5. Diferencia  $\Delta$  *calibration slope* sin validación interna

MCAR

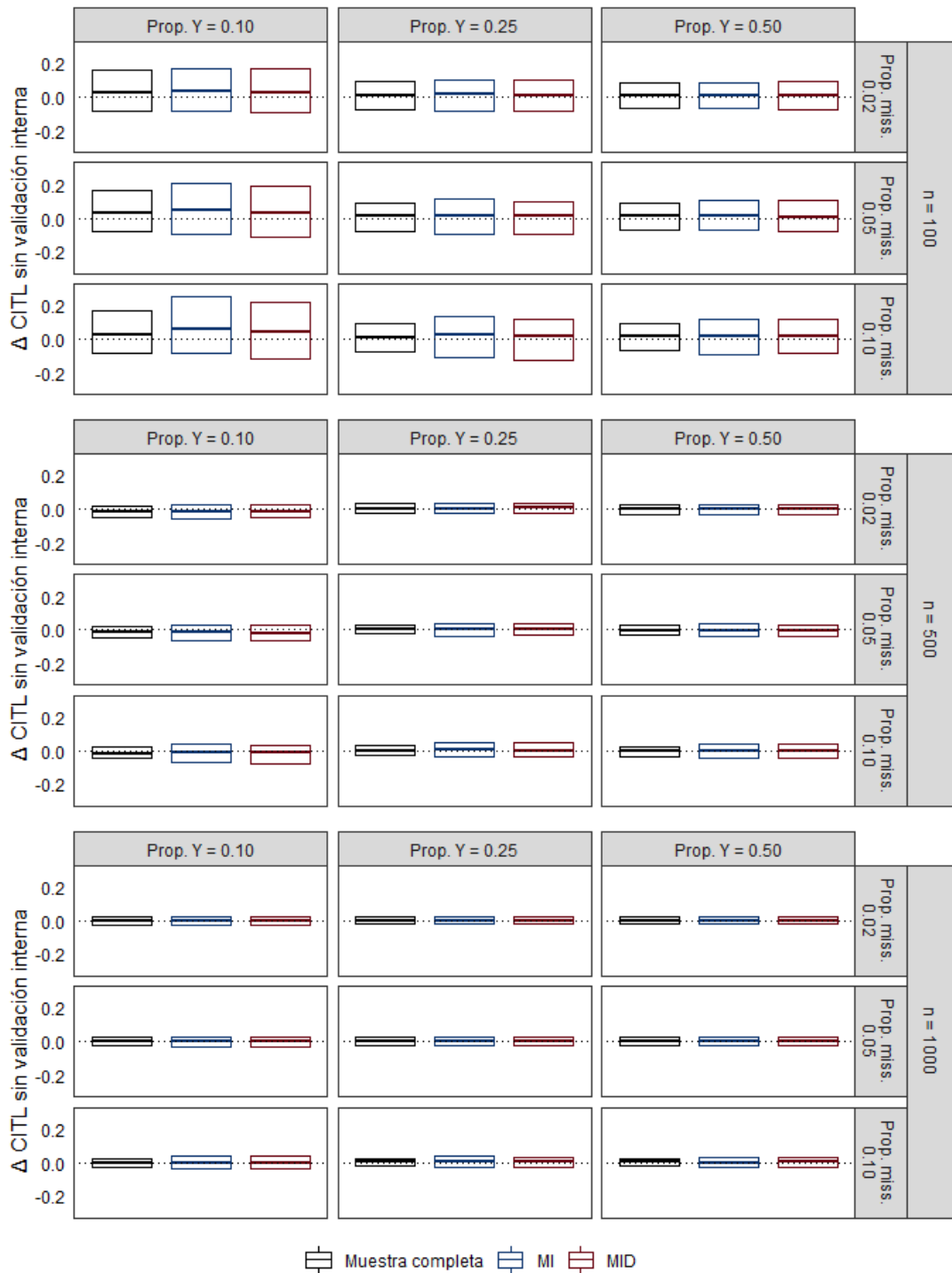


MAR

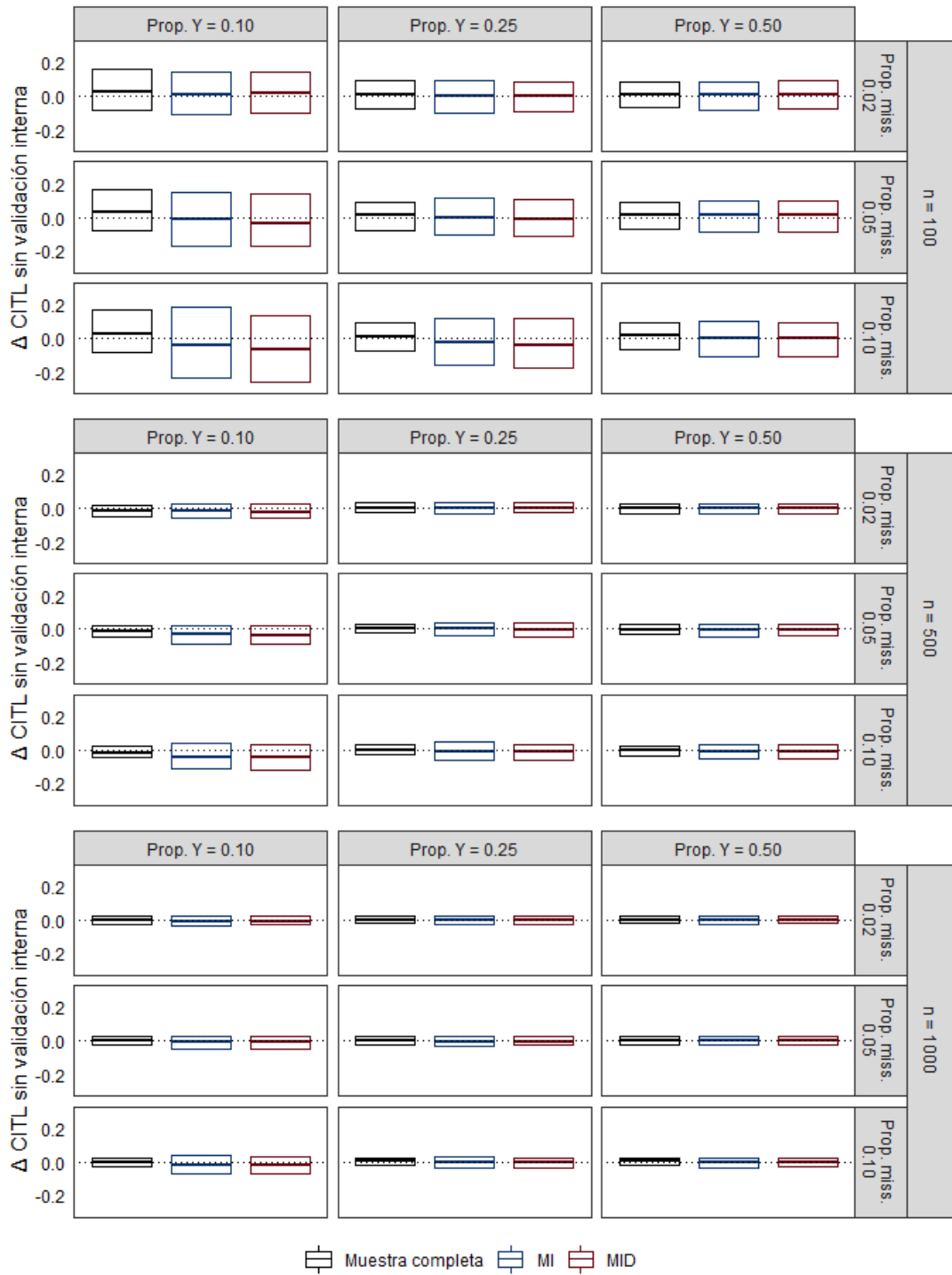


Anexo 6. Diferencia  $\Delta$  CITL sin validación interna

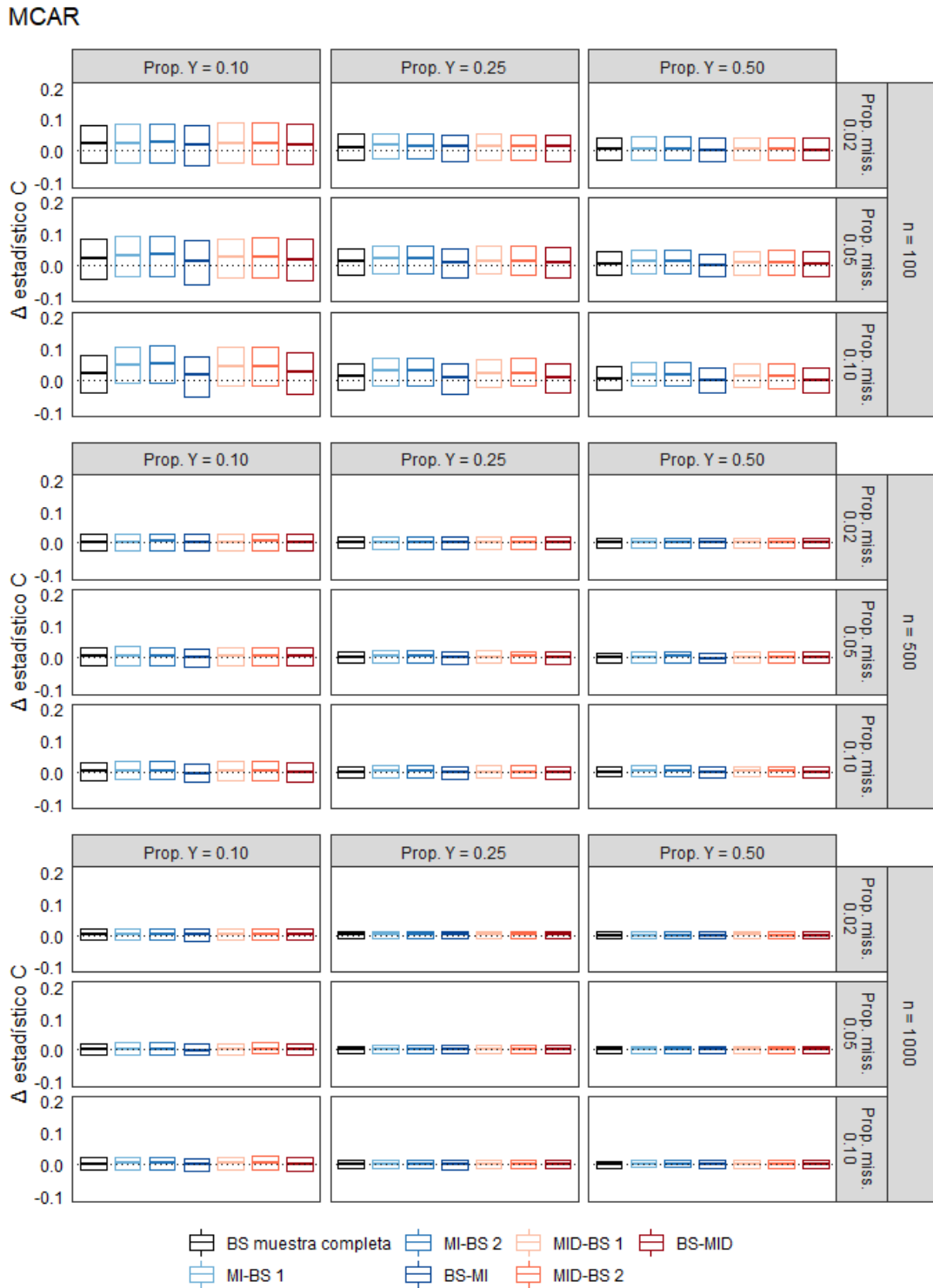
MCAR



MAR



Anexo 7. Diferencia  $\Delta$  estadístico C por estrategias: gráficos de caja bajo MCAR



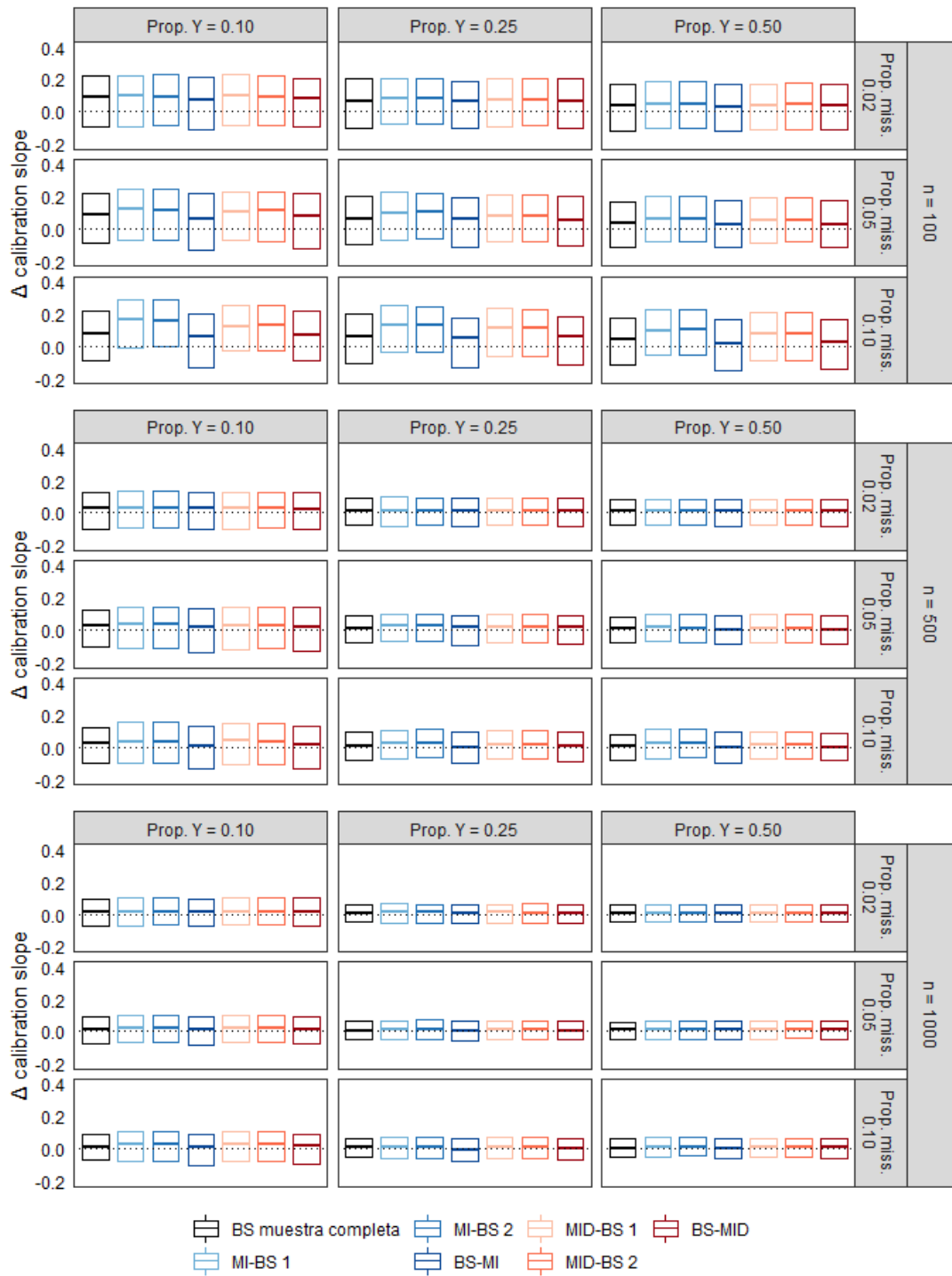
**Anexo 8. Diferencia  $\Delta$  estadístico C por estrategias: valores medios bajo MCAR**

Valores medios diferencia  $\Delta$  estadístico C (MCAR)

n	propY	propmiss	EPV	MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID
100	0.10	0.02	2	0.0217	0.0221	0.0151	0.0206	0.0203	0.0168
		0.05		0.0298	0.0299	0.0111	0.0262	0.0255	0.0158
		0.10		0.0488	0.0489	0.0115	0.0398	0.0399	0.0196
	0.25	0.02	5	0.0107	0.0109	0.0061	0.0093	0.0096	0.0069
		0.05		0.0173	0.0172	0.0056	0.0136	0.0135	0.0069
		0.10		0.0246	0.0246	0.0003	0.0181	0.0181	0.0037
	0.50	0.02	10	0.0048	0.0052	0.0013	0.0039	0.0038	0.0016
		0.05		0.0103	0.0103	0.0006	0.0073	0.0071	0.0018
		0.10		0.0192	0.0196	-0.0024	0.0124	0.0122	-0.0006
500	0.10	0.02	10	0.0017	0.0019	0.0001	0.0015	0.0016	0.0005
		0.05		0.0033	0.0033	-0.0010	0.0022	0.0021	-0.0004
		0.10		0.0062	0.0061	-0.0042	0.0041	0.0040	-0.0017
	0.25	0.02	25	0.0005	0.0004	-0.0008	0.0004	0.0002	-0.0003
		0.05		0.0022	0.0021	-0.0008	0.0018	0.0018	0.0002
		0.10		0.0036	0.0036	-0.0019	0.0020	0.0020	-0.0011
	0.50	0.02	50	0.0011	0.0010	0.0002	0.0006	0.0006	0.0002
		0.05		0.0018	0.0019	-0.0004	0.0012	0.0013	0.0000
		0.10		0.0041	0.0040	-0.0005	0.0023	0.0023	-0.0001
1000	0.10	0.02	20	0.0024	0.0024	0.0015	0.0023	0.0021	0.0016
		0.05		0.0028	0.0026	0.0002	0.0021	0.0023	0.0009
		0.10		0.0043	0.0043	-0.0010	0.0035	0.0034	0.0006
	0.25	0.02	50	0.0007	0.0007	0.0004	0.0006	0.0005	0.0003
		0.05		0.0021	0.0021	0.0008	0.0017	0.0018	0.0010
		0.10		0.0015	0.0015	-0.0017	0.0007	0.0009	-0.0008
	0.50	0.02	100	0.0004	0.0004	0.0000	0.0003	0.0002	-0.0001
		0.05		0.0011	0.0012	0.0003	0.0009	0.0008	0.0004
		0.10		0.0020	0.0020	-0.0003	0.0014	0.0014	0.0000

Anexo 9. Diferencia  $\Delta$  calibration slope por estrategias: gráficos de caja bajo MCAR

MCAR



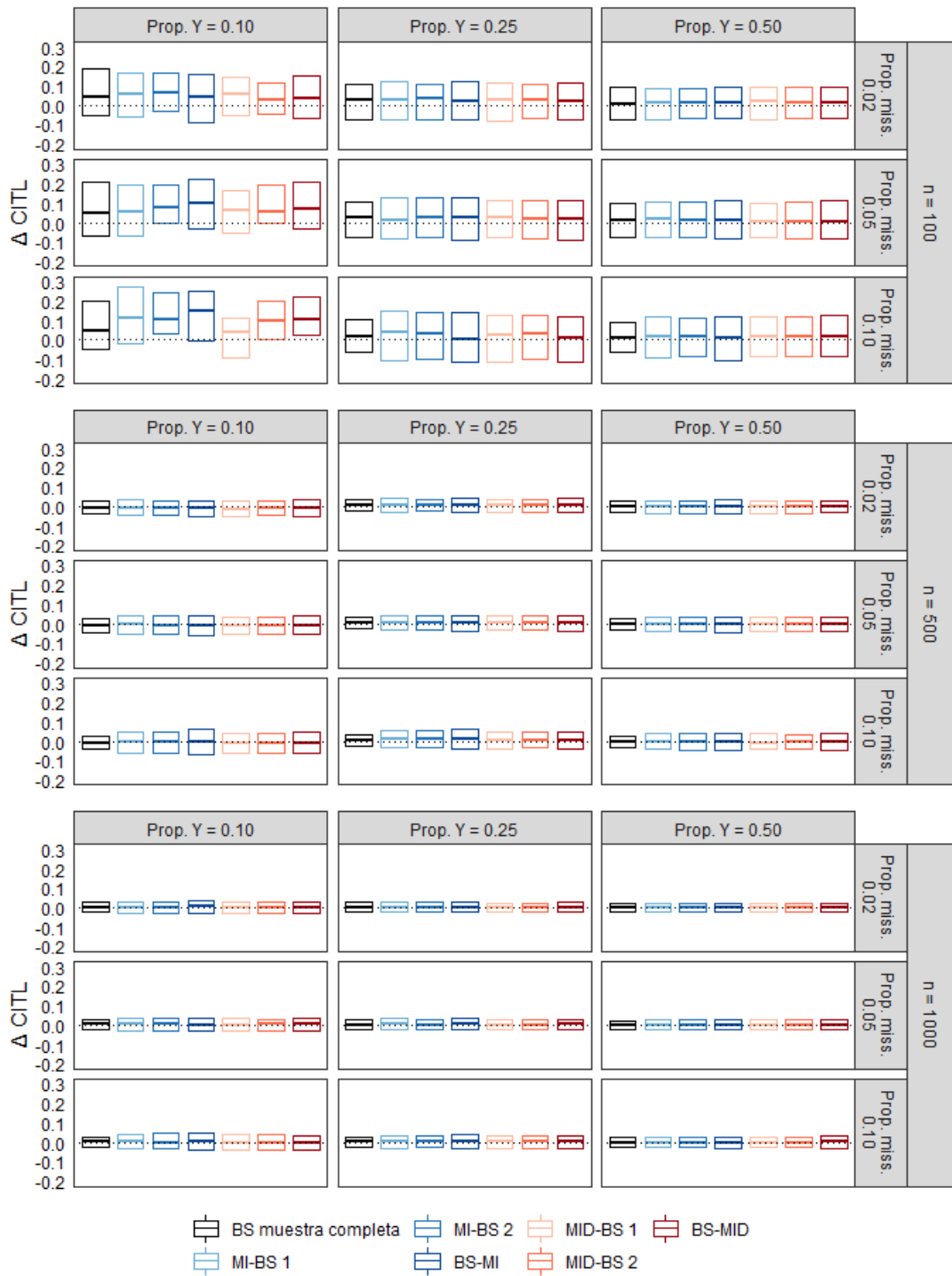
**Anexo 10. Diferencia  $\Delta$  calibration slope por estrategias: valores medios bajo MCAR**

Valores medios diferencia  $\Delta$  calibration slope (MCAR)

n	propY	propmiss	EPV	MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID
100	0.10	0.02	2	0.0375	0.0385	0.0184	0.0343	0.0329	0.0228
		0.05		0.0568	0.0567	0.0034	0.0467	0.0440	0.0169
		0.10		0.1118	0.1121	0.0108	0.0797	0.0804	0.0280
	0.25	0.02	5	0.0261	0.0264	0.0083	0.0207	0.0221	0.0118
		0.05		0.0513	0.0510	0.0068	0.0382	0.0371	0.0117
		0.10		0.0717	0.0719	-0.0168	0.0508	0.0507	-0.0022
	0.50	0.02	10	0.0109	0.0127	-0.0034	0.0080	0.0076	-0.0020
		0.05		0.0336	0.0326	-0.0055	0.0205	0.0198	-0.0014
		0.10		0.0661	0.0675	-0.0182	0.0417	0.0410	-0.0104
500	0.10	0.02	10	-0.0024	-0.0021	-0.0090	-0.0036	-0.0029	-0.0071
		0.05		0.0001	-0.0001	-0.0177	-0.0036	-0.0040	-0.0148
		0.10		0.0118	0.0113	-0.0301	0.0060	0.0056	-0.0178
	0.25	0.02	25	0.0009	0.0004	-0.0046	0.0005	-0.0003	-0.0026
		0.05		0.0096	0.0089	-0.0035	0.0082	0.0080	0.0011
		0.10		0.0123	0.0121	-0.0115	0.0058	0.0063	-0.0074
	0.50	0.02	50	0.0027	0.0028	-0.0011	0.0010	0.0008	-0.0010
		0.05		0.0050	0.0054	-0.0051	0.0027	0.0031	-0.0028
		0.10		0.0156	0.0151	-0.0054	0.0092	0.0090	-0.0029
1000	0.10	0.02	20	0.0091	0.0090	0.0054	0.0086	0.0079	0.0058
		0.05		0.0112	0.0104	0.0006	0.0089	0.0097	0.0042
		0.10		0.0125	0.0127	-0.0095	0.0105	0.0103	-0.0017
	0.25	0.02	50	0.0030	0.0036	0.0019	0.0033	0.0026	0.0017
		0.05		0.0086	0.0085	0.0025	0.0069	0.0074	0.0036
		0.10		0.0059	0.0061	-0.0077	0.0034	0.0040	-0.0036
	0.50	0.02	100	0.0038	0.0037	0.0018	0.0031	0.0028	0.0016
		0.05		0.0072	0.0076	0.0036	0.0062	0.0060	0.0040
		0.10		0.0097	0.0099	-0.0007	0.0069	0.0069	0.0006

Anexo 11. Diferencia  $\Delta$  CITL por estrategias: gráficos de caja bajo MCAR

MCAR



**Anexo 12. Diferencia  $\Delta$  CITL por estrategias: valores medios bajo MCAR**

Valores medios diferencia  $\Delta$  CITL (MCAR)

n	propY	propmiss	EPV	MI-BS 1	MI-BS 2	BS-MI	MID-BS 1	MID-BS 2	BS-MID
100	0.10	0.02	2	0.0677	0.0717	0.0526	0.0625	0.0366	0.0550
		0.05		0.0705	0.0973	0.1045	0.0723	0.1063	0.0893
		0.10		0.1557	0.1268	0.1401	0.0233	0.1041	0.1129
	0.25	0.02	5	0.0245	0.0238	0.0212	0.0225	0.0228	0.0211
		0.05		0.0214	0.0237	0.0180	0.0170	0.0211	0.0117
		0.10		0.0279	0.0273	0.0064	0.0185	0.0228	0.0073
	0.50	0.02	10	0.0112	0.0135	0.0145	0.0145	0.0132	0.0144
		0.05		0.0163	0.0137	0.0153	0.0123	0.0128	0.0146
		0.10		0.0177	0.0162	0.0129	0.0203	0.0210	0.0167
500	0.10	0.02	10	-0.0078	-0.0087	-0.0095	-0.0109	-0.0083	-0.0085
		0.05		-0.0068	-0.0071	-0.0074	-0.0102	-0.0105	-0.0104
		0.10		-0.0063	-0.0057	-0.0059	-0.0116	-0.0120	-0.0116
	0.25	0.02	25	0.0063	0.0061	0.0063	0.0041	0.0059	0.0056
		0.05		0.0042	0.0042	0.0040	0.0030	0.0035	0.0030
		0.10		0.0098	0.0094	0.0105	0.0058	0.0062	0.0059
	0.50	0.02	50	-0.0026	-0.0017	-0.0009	-0.0019	-0.0021	-0.0014
		0.05		-0.0033	-0.0029	-0.0036	-0.0021	-0.0028	-0.0022
		0.10		-0.0017	-0.0017	-0.0006	-0.0031	-0.0025	-0.0023
1000	0.10	0.02	20	0.0029	0.0033	0.0048	0.0026	0.0032	0.0041
		0.05		0.0022	0.0026	0.0023	0.0020	0.0019	0.0014
		0.10		0.0036	0.0041	0.0051	-0.0008	-0.0009	-0.0006
	0.25	0.02	50	0.0033	0.0036	0.0041	0.0034	0.0033	0.0037
		0.05		0.0049	0.0048	0.0054	0.0036	0.0038	0.0033
		0.10		0.0058	0.0053	0.0047	0.0040	0.0043	0.0041
	0.50	0.02	100	0.0013	0.0012	0.0015	0.0012	0.0010	0.0011
		0.05		0.0018	0.0017	0.0013	0.0017	0.0019	0.0016
		0.10		0.0011	0.0008	0.0000	0.0015	0.0015	0.0012

**Anexo 13. Diferencia  $\Delta$  estadístico C por estrategias: estadísticos descriptivos bajo MCAR**

Estadísticos descriptivos diferencia $\Delta$ estadístico C (MCAR)											
n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.10	0.02	MI-BS 1	500	0.0217	0.0895	0.0248	-0.0379	0.0842	-0.1658	0.1755
100	0.10	0.02	MI-BS 2	500	0.0221	0.0890	0.0259	-0.0360	0.0836	-0.1625	0.1772
100	0.10	0.02	BS-MI	500	0.0151	0.0915	0.0179	-0.0457	0.0805	-0.1682	0.1718
100	0.10	0.02	MID-BS 1	500	0.0206	0.0904	0.0226	-0.0388	0.0869	-0.1679	0.1775
100	0.10	0.02	MID-BS 2	500	0.0203	0.0901	0.0211	-0.0402	0.0873	-0.1673	0.1710
100	0.10	0.02	BS-MID	500	0.0168	0.0911	0.0207	-0.0438	0.0843	-0.1658	0.1708
100	0.10	0.05	MI-BS 1	500	0.0297	0.0906	0.0328	-0.0334	0.0930	-0.1432	0.1978
100	0.10	0.05	MI-BS 2	500	0.0299	0.0903	0.0342	-0.0339	0.0922	-0.1412	0.2012
100	0.10	0.05	BS-MI	500	0.0111	0.0957	0.0152	-0.0584	0.0808	-0.1737	0.1817
100	0.10	0.05	MID-BS 1	500	0.0262	0.0927	0.0285	-0.0385	0.0855	-0.1495	0.2011
100	0.10	0.05	MID-BS 2	500	0.0255	0.0933	0.0259	-0.0401	0.0859	-0.1530	0.2032
100	0.10	0.05	BS-MID	500	0.0158	0.0960	0.0189	-0.0485	0.0825	-0.1661	0.1920
100	0.10	0.10	MI-BS 1	500	0.0488	0.0867	0.0499	-0.0079	0.1052	-0.1275	0.2030
100	0.10	0.10	MI-BS 2	500	0.0489	0.0869	0.0541	-0.0081	0.1072	-0.1215	0.2063
100	0.10	0.10	BS-MI	500	0.0115	0.0970	0.0156	-0.0544	0.0760	-0.1777	0.1891
100	0.10	0.10	MID-BS 1	500	0.0398	0.0926	0.0427	-0.0156	0.1057	-0.1559	0.2055
100	0.10	0.10	MID-BS 2	500	0.0399	0.0927	0.0439	-0.0194	0.1063	-0.1500	0.2076
100	0.10	0.10	BS-MID	500	0.0196	0.0980	0.0250	-0.0452	0.0875	-0.1773	0.1974
100	0.25	0.02	MI-BS 1	500	0.0107	0.0644	0.0168	-0.0264	0.0533	-0.1361	0.1230
100	0.25	0.02	MI-BS 2	500	0.0109	0.0641	0.0153	-0.0268	0.0528	-0.1369	0.1223
100	0.25	0.02	BS-MI	500	0.0060	0.0655	0.0128	-0.0325	0.0479	-0.1495	0.1205
100	0.25	0.02	MID-BS 1	500	0.0093	0.0649	0.0148	-0.0291	0.0546	-0.1385	0.1232
100	0.25	0.02	MID-BS 2	500	0.0096	0.0644	0.0148	-0.0283	0.0515	-0.1376	0.1213
100	0.25	0.02	BS-MID	500	0.0069	0.0653	0.0138	-0.0336	0.0495	-0.1387	0.1201
100	0.25	0.05	MI-BS 1	500	0.0173	0.0634	0.0207	-0.0235	0.0603	-0.1288	0.1366
100	0.25	0.05	MI-BS 2	500	0.0172	0.0633	0.0210	-0.0243	0.0601	-0.1273	0.1349
100	0.25	0.05	BS-MI	500	0.0056	0.0662	0.0092	-0.0395	0.0521	-0.1447	0.1283
100	0.25	0.05	MID-BS 1	500	0.0136	0.0653	0.0157	-0.0277	0.0599	-0.1293	0.1291
100	0.25	0.05	MID-BS 2	500	0.0135	0.0651	0.0157	-0.0286	0.0596	-0.1298	0.1315
100	0.25	0.05	BS-MID	500	0.0069	0.0667	0.0107	-0.0373	0.0555	-0.1382	0.1282
100	0.25	0.10	MI-BS 1	500	0.0246	0.0661	0.0300	-0.0168	0.0706	-0.1258	0.1421
100	0.25	0.10	MI-BS 2	500	0.0246	0.0657	0.0310	-0.0167	0.0708	-0.1204	0.1414
100	0.25	0.10	BS-MI	500	0.0003	0.0719	0.0066	-0.0426	0.0505	-0.1704	0.1256
100	0.25	0.10	MID-BS 1	500	0.0181	0.0686	0.0230	-0.0213	0.0674	-0.1372	0.1383
100	0.25	0.10	MID-BS 2	500	0.0181	0.0685	0.0224	-0.0190	0.0680	-0.1368	0.1361
100	0.25	0.10	BS-MID	500	0.0037	0.0715	0.0079	-0.0389	0.0530	-0.1571	0.1300
100	0.50	0.02	MI-BS 1	500	0.0048	0.0552	0.0044	-0.0295	0.0444	-0.1151	0.1028

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.50	0.02	MI-BS 2	500	0.0052	0.0550	0.0059	-0.0285	0.0447	-0.1112	0.1012
100	0.50	0.02	BS-MI	500	0.0013	0.0561	-0.0004	-0.0327	0.0425	-0.1199	0.1018
100	0.50	0.02	MID-BS 1	500	0.0039	0.0555	0.0055	-0.0304	0.0424	-0.1117	0.1012
100	0.50	0.02	MID-BS 2	500	0.0038	0.0554	0.0042	-0.0308	0.0426	-0.1156	0.0999
100	0.50	0.02	BS-MID	500	0.0016	0.0559	0.0021	-0.0311	0.0412	-0.1192	0.1017
100	0.50	0.05	MI-BS 1	500	0.0103	0.0556	0.0131	-0.0254	0.0479	-0.1136	0.1132
100	0.50	0.05	MI-BS 2	500	0.0103	0.0554	0.0131	-0.0251	0.0482	-0.1145	0.1106
100	0.50	0.05	BS-MI	500	0.0006	0.0579	0.0020	-0.0362	0.0377	-0.1303	0.1066
100	0.50	0.05	MID-BS 1	500	0.0073	0.0568	0.0102	-0.0319	0.0460	-0.1206	0.1067
100	0.50	0.05	MID-BS 2	500	0.0071	0.0568	0.0111	-0.0294	0.0465	-0.1188	0.1109
100	0.50	0.05	BS-MID	500	0.0018	0.0582	0.0045	-0.0341	0.0425	-0.1239	0.1059
100	0.50	0.10	MI-BS 1	500	0.0192	0.0568	0.0194	-0.0160	0.0572	-0.0982	0.1237
100	0.50	0.10	MI-BS 2	500	0.0196	0.0568	0.0191	-0.0159	0.0575	-0.0974	0.1284
100	0.50	0.10	BS-MI	500	-0.0024	0.0622	-0.0021	-0.0409	0.0399	-0.1303	0.1154
100	0.50	0.10	MID-BS 1	500	0.0124	0.0595	0.0121	-0.0241	0.0539	-0.1128	0.1266
100	0.50	0.10	MID-BS 2	500	0.0122	0.0593	0.0121	-0.0256	0.0531	-0.1124	0.1265
100	0.50	0.10	BS-MID	500	-0.0006	0.0627	-0.0005	-0.0410	0.0408	-0.1251	0.1175
500	0.10	0.02	MI-BS 1	500	0.0017	0.0402	0.0038	-0.0229	0.0300	-0.0829	0.0708
500	0.10	0.02	MI-BS 2	500	0.0019	0.0400	0.0053	-0.0241	0.0300	-0.0825	0.0687
500	0.10	0.02	BS-MI	500	0.0001	0.0404	0.0024	-0.0255	0.0272	-0.0852	0.0721
500	0.10	0.02	MID-BS 1	500	0.0015	0.0404	0.0039	-0.0251	0.0296	-0.0811	0.0722
500	0.10	0.02	MID-BS 2	500	0.0016	0.0403	0.0053	-0.0252	0.0292	-0.0797	0.0723
500	0.10	0.02	BS-MID	500	0.0005	0.0404	0.0033	-0.0249	0.0280	-0.0823	0.0719
500	0.10	0.05	MI-BS 1	500	0.0033	0.0413	0.0060	-0.0243	0.0341	-0.0804	0.0743
500	0.10	0.05	MI-BS 2	500	0.0033	0.0410	0.0061	-0.0240	0.0332	-0.0782	0.0746
500	0.10	0.05	BS-MI	500	-0.0010	0.0430	0.0026	-0.0286	0.0289	-0.0897	0.0751
500	0.10	0.05	MID-BS 1	500	0.0022	0.0418	0.0057	-0.0268	0.0323	-0.0825	0.0737
500	0.10	0.05	MID-BS 2	500	0.0021	0.0417	0.0060	-0.0259	0.0314	-0.0805	0.0721
500	0.10	0.05	BS-MID	500	-0.0004	0.0427	0.0042	-0.0264	0.0306	-0.0877	0.0722
500	0.10	0.10	MI-BS 1	500	0.0062	0.0440	0.0060	-0.0208	0.0363	-0.0875	0.0871
500	0.10	0.10	MI-BS 2	500	0.0061	0.0437	0.0067	-0.0193	0.0363	-0.0848	0.0871
500	0.10	0.10	BS-MI	500	-0.0042	0.0476	-0.0022	-0.0323	0.0281	-0.1076	0.0787
500	0.10	0.10	MID-BS 1	500	0.0041	0.0445	0.0059	-0.0241	0.0349	-0.0850	0.0875
500	0.10	0.10	MID-BS 2	500	0.0040	0.0444	0.0056	-0.0243	0.0345	-0.0873	0.0855
500	0.10	0.10	BS-MID	500	-0.0017	0.0466	0.0006	-0.0304	0.0312	-0.0991	0.0828
500	0.25	0.02	MI-BS 1	500	0.0004	0.0270	0.0010	-0.0185	0.0202	-0.0545	0.0504
500	0.25	0.02	MI-BS 2	500	0.0004	0.0269	0.0012	-0.0185	0.0194	-0.0538	0.0494
500	0.25	0.02	BS-MI	500	-0.0008	0.0273	-0.0001	-0.0206	0.0185	-0.0559	0.0498
500	0.25	0.02	MID-BS 1	500	0.0004	0.0272	0.0014	-0.0183	0.0193	-0.0542	0.0481

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
500	0.25	0.02	MID-BS 2	500	0.0002	0.0271	0.0012	-0.0185	0.0195	-0.0541	0.0484
500	0.25	0.02	BS-MID	500	-0.0003	0.0272	0.0006	-0.0197	0.0199	-0.0574	0.0505
500	0.25	0.05	MI-BS 1	500	0.0022	0.0276	0.0038	-0.0177	0.0211	-0.0530	0.0518
500	0.25	0.05	MI-BS 2	500	0.0021	0.0276	0.0039	-0.0178	0.0209	-0.0541	0.0538
500	0.25	0.05	BS-MI	500	-0.0008	0.0285	0.0006	-0.0219	0.0188	-0.0590	0.0535
500	0.25	0.05	MID-BS 1	500	0.0018	0.0276	0.0033	-0.0176	0.0214	-0.0541	0.0541
500	0.25	0.05	MID-BS 2	500	0.0018	0.0275	0.0036	-0.0184	0.0206	-0.0543	0.0526
500	0.25	0.05	BS-MID	500	0.0002	0.0281	0.0008	-0.0201	0.0196	-0.0560	0.0552
500	0.25	0.10	MI-BS 1	500	0.0036	0.0288	0.0041	-0.0150	0.0232	-0.0551	0.0584
500	0.25	0.10	MI-BS 2	500	0.0036	0.0287	0.0041	-0.0149	0.0237	-0.0550	0.0573
500	0.25	0.10	BS-MI	500	-0.0019	0.0307	-0.0012	-0.0221	0.0189	-0.0650	0.0578
500	0.25	0.10	MID-BS 1	500	0.0020	0.0292	0.0023	-0.0162	0.0222	-0.0592	0.0546
500	0.25	0.10	MID-BS 2	500	0.0020	0.0292	0.0016	-0.0160	0.0208	-0.0583	0.0548
500	0.25	0.10	BS-MID	500	-0.0011	0.0303	-0.0007	-0.0212	0.0185	-0.0663	0.0553
500	0.50	0.02	MI-BS 1	500	0.0011	0.0227	0.0012	-0.0144	0.0169	-0.0442	0.0431
500	0.50	0.02	MI-BS 2	500	0.0010	0.0227	0.0012	-0.0141	0.0162	-0.0441	0.0423
500	0.50	0.02	BS-MI	500	0.0002	0.0231	0.0011	-0.0159	0.0163	-0.0453	0.0440
500	0.50	0.02	MID-BS 1	500	0.0006	0.0227	0.0009	-0.0149	0.0167	-0.0437	0.0420
500	0.50	0.02	MID-BS 2	500	0.0006	0.0227	0.0012	-0.0139	0.0161	-0.0439	0.0422
500	0.50	0.02	BS-MID	500	0.0002	0.0231	0.0006	-0.0147	0.0158	-0.0446	0.0430
500	0.50	0.05	MI-BS 1	500	0.0018	0.0242	0.0027	-0.0143	0.0177	-0.0436	0.0474
500	0.50	0.05	MI-BS 2	500	0.0019	0.0242	0.0034	-0.0149	0.0180	-0.0437	0.0452
500	0.50	0.05	BS-MI	500	-0.0004	0.0252	-0.0013	-0.0176	0.0163	-0.0490	0.0446
500	0.50	0.05	MID-BS 1	500	0.0012	0.0246	0.0022	-0.0155	0.0172	-0.0444	0.0468
500	0.50	0.05	MID-BS 2	500	0.0013	0.0246	0.0026	-0.0151	0.0180	-0.0457	0.0465
500	0.50	0.05	BS-MID	500	0.0000	0.0253	0.0004	-0.0167	0.0167	-0.0496	0.0460
500	0.50	0.10	MI-BS 1	500	0.0041	0.0248	0.0057	-0.0132	0.0214	-0.0433	0.0521
500	0.50	0.10	MI-BS 2	500	0.0040	0.0248	0.0051	-0.0134	0.0216	-0.0420	0.0512
500	0.50	0.10	BS-MI	500	-0.0005	0.0266	-0.0009	-0.0183	0.0185	-0.0513	0.0507
500	0.50	0.10	MID-BS 1	500	0.0023	0.0247	0.0036	-0.0144	0.0198	-0.0443	0.0481
500	0.50	0.10	MID-BS 2	500	0.0023	0.0247	0.0035	-0.0145	0.0194	-0.0452	0.0483
500	0.50	0.10	BS-MID	500	-0.0001	0.0258	0.0010	-0.0172	0.0186	-0.0513	0.0469
1000	0.10	0.02	MI-BS 1	500	0.0024	0.0265	0.0022	-0.0160	0.0221	-0.0493	0.0515
1000	0.10	0.02	MI-BS 2	500	0.0024	0.0263	0.0024	-0.0162	0.0213	-0.0477	0.0502
1000	0.10	0.02	BS-MI	500	0.0015	0.0270	0.0019	-0.0177	0.0207	-0.0513	0.0525
1000	0.10	0.02	MID-BS 1	500	0.0023	0.0264	0.0021	-0.0151	0.0210	-0.0504	0.0501
1000	0.10	0.02	MID-BS 2	500	0.0021	0.0262	0.0022	-0.0156	0.0210	-0.0492	0.0489
1000	0.10	0.02	BS-MID	500	0.0016	0.0266	0.0019	-0.0166	0.0211	-0.0506	0.0503
1000	0.10	0.05	MI-BS 1	500	0.0028	0.0270	0.0021	-0.0143	0.0225	-0.0489	0.0518

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.10	0.05	MI-BS 2	500	0.0026	0.0268	0.0024	-0.0141	0.0217	-0.0492	0.0517
1000	0.10	0.05	BS-MI	500	0.0002	0.0282	-0.0010	-0.0178	0.0205	-0.0552	0.0542
1000	0.10	0.05	MID-BS 1	500	0.0021	0.0271	0.0016	-0.0150	0.0208	-0.0512	0.0515
1000	0.10	0.05	MID-BS 2	500	0.0023	0.0269	0.0020	-0.0134	0.0218	-0.0515	0.0517
1000	0.10	0.05	BS-MID	500	0.0009	0.0277	0.0000	-0.0147	0.0197	-0.0558	0.0519
1000	0.10	0.10	MI-BS 1	500	0.0043	0.0290	0.0046	-0.0153	0.0244	-0.0550	0.0589
1000	0.10	0.10	MI-BS 2	500	0.0043	0.0289	0.0043	-0.0151	0.0250	-0.0538	0.0575
1000	0.10	0.10	BS-MI	500	-0.0010	0.0308	0.0003	-0.0230	0.0204	-0.0603	0.0566
1000	0.10	0.10	MID-BS 1	500	0.0035	0.0294	0.0034	-0.0157	0.0243	-0.0537	0.0578
1000	0.10	0.10	MID-BS 2	500	0.0034	0.0294	0.0035	-0.0158	0.0252	-0.0543	0.0574
1000	0.10	0.10	BS-MID	500	0.0006	0.0306	0.0013	-0.0195	0.0218	-0.0594	0.0571
1000	0.25	0.02	MI-BS 1	500	0.0006	0.0174	0.0010	-0.0109	0.0125	-0.0334	0.0351
1000	0.25	0.02	MI-BS 2	500	0.0007	0.0172	0.0016	-0.0106	0.0122	-0.0333	0.0345
1000	0.25	0.02	BS-MI	500	0.0004	0.0176	0.0013	-0.0117	0.0122	-0.0322	0.0328
1000	0.25	0.02	MID-BS 1	500	0.0006	0.0175	0.0013	-0.0106	0.0120	-0.0338	0.0349
1000	0.25	0.02	MID-BS 2	500	0.0005	0.0173	0.0010	-0.0108	0.0121	-0.0329	0.0351
1000	0.25	0.02	BS-MID	500	0.0003	0.0176	0.0008	-0.0113	0.0115	-0.0325	0.0335
1000	0.25	0.05	MI-BS 1	500	0.0021	0.0184	0.0029	-0.0100	0.0149	-0.0347	0.0361
1000	0.25	0.05	MI-BS 2	500	0.0021	0.0182	0.0026	-0.0098	0.0149	-0.0342	0.0352
1000	0.25	0.05	BS-MI	500	0.0008	0.0188	0.0009	-0.0117	0.0137	-0.0363	0.0358
1000	0.25	0.05	MID-BS 1	500	0.0017	0.0184	0.0022	-0.0102	0.0137	-0.0351	0.0349
1000	0.25	0.05	MID-BS 2	500	0.0018	0.0183	0.0024	-0.0108	0.0139	-0.0355	0.0351
1000	0.25	0.05	BS-MID	500	0.0010	0.0187	0.0015	-0.0114	0.0135	-0.0367	0.0365
1000	0.25	0.10	MI-BS 1	500	0.0015	0.0189	0.0026	-0.0111	0.0140	-0.0363	0.0378
1000	0.25	0.10	MI-BS 2	500	0.0015	0.0190	0.0021	-0.0116	0.0146	-0.0363	0.0387
1000	0.25	0.10	BS-MI	500	-0.0017	0.0204	-0.0009	-0.0156	0.0131	-0.0408	0.0387
1000	0.25	0.10	MID-BS 1	500	0.0007	0.0191	0.0008	-0.0126	0.0136	-0.0390	0.0359
1000	0.25	0.10	MID-BS 2	500	0.0009	0.0190	0.0016	-0.0116	0.0137	-0.0375	0.0359
1000	0.25	0.10	BS-MID	500	-0.0008	0.0199	-0.0003	-0.0140	0.0127	-0.0410	0.0366
1000	0.50	0.02	MI-BS 1	500	0.0004	0.0155	0.0003	-0.0109	0.0118	-0.0313	0.0285
1000	0.50	0.02	MI-BS 2	500	0.0004	0.0154	-0.0002	-0.0102	0.0112	-0.0308	0.0294
1000	0.50	0.02	BS-MI	500	0.0000	0.0156	-0.0002	-0.0105	0.0109	-0.0319	0.0291
1000	0.50	0.02	MID-BS 1	500	0.0003	0.0154	0.0006	-0.0106	0.0109	-0.0304	0.0293
1000	0.50	0.02	MID-BS 2	500	0.0002	0.0154	-0.0003	-0.0105	0.0108	-0.0311	0.0289
1000	0.50	0.02	BS-MID	500	-0.0001	0.0155	-0.0003	-0.0108	0.0107	-0.0317	0.0288
1000	0.50	0.05	MI-BS 1	500	0.0011	0.0159	0.0017	-0.0096	0.0117	-0.0312	0.0319
1000	0.50	0.05	MI-BS 2	500	0.0012	0.0159	0.0017	-0.0100	0.0118	-0.0304	0.0310
1000	0.50	0.05	BS-MI	500	0.0003	0.0165	0.0009	-0.0111	0.0120	-0.0337	0.0314
1000	0.50	0.05	MID-BS 1	500	0.0009	0.0159	0.0013	-0.0103	0.0111	-0.0302	0.0310

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MCAR)

<b>n</b>	<b>propY</b>	<b>propmiss</b>	<b>Estrategia</b>	<b>NSIM</b>	<b>Media</b>	<b>SD</b>	<b>Mediana</b>	<b>Q1</b>	<b>Q3</b>	<b>P2.5</b>	<b>P97.5</b>
1000	0.50	0.05	MID-BS 2	500	0.0008	0.0159	0.0013	-0.0105	0.0113	-0.0310	0.0304
1000	0.50	0.05	BS-MID	500	0.0004	0.0163	0.0006	-0.0109	0.0115	-0.0317	0.0306
1000	0.50	0.10	MI-BS 1	500	0.0020	0.0165	0.0023	-0.0096	0.0133	-0.0287	0.0326
1000	0.50	0.10	MI-BS 2	500	0.0020	0.0166	0.0022	-0.0098	0.0136	-0.0289	0.0316
1000	0.50	0.10	BS-MI	500	-0.0003	0.0177	-0.0008	-0.0128	0.0122	-0.0314	0.0322
1000	0.50	0.10	MID-BS 1	500	0.0014	0.0166	0.0025	-0.0109	0.0129	-0.0299	0.0317
1000	0.50	0.10	MID-BS 2	500	0.0014	0.0165	0.0017	-0.0104	0.0129	-0.0296	0.0317
1000	0.50	0.10	BS-MID	500	0.0000	0.0170	-0.0005	-0.0118	0.0123	-0.0310	0.0329

**Anexo 14. Diferencia  $\Delta$  estadístico C por estrategias: estadísticos descriptivos bajo MAR**

Estadísticos descriptivos diferencia $\Delta$ estadístico C (MAR)											
n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.10	0.02	MI-BS 1	500	0.0308	0.0953	0.0403	-0.0332	0.1019	-0.1724	0.1902
100	0.10	0.02	MI-BS 2	500	0.0306	0.0953	0.0393	-0.0343	0.1004	-0.1714	0.1920
100	0.10	0.02	BS-MI	500	0.0119	0.0992	0.0190	-0.0569	0.0858	-0.1960	0.1795
100	0.10	0.02	MID-BS 1	500	0.0263	0.0968	0.0351	-0.0414	0.0944	-0.1770	0.1914
100	0.10	0.02	MID-BS 2	500	0.0259	0.0970	0.0364	-0.0423	0.0942	-0.1791	0.1873
100	0.10	0.02	BS-MID	500	0.0123	0.1000	0.0214	-0.0581	0.0823	-0.1918	0.1783
100	0.10	0.05	MI-BS 1	500	0.0503	0.0964	0.0529	-0.0169	0.1179	-0.1309	0.2358
100	0.10	0.05	MI-BS 2	500	0.0502	0.0967	0.0550	-0.0137	0.1166	-0.1336	0.2432
100	0.10	0.05	BS-MI	500	0.0072	0.1063	0.0118	-0.0648	0.0790	-0.1914	0.2128
100	0.10	0.05	MID-BS 1	500	0.0412	0.1009	0.0488	-0.0254	0.1083	-0.1641	0.2358
100	0.10	0.05	MID-BS 2	500	0.0412	0.1006	0.0488	-0.0268	0.1077	-0.1599	0.2380
100	0.10	0.05	BS-MID	500	0.0094	0.1076	0.0199	-0.0653	0.0793	-0.2020	0.2151
100	0.10	0.10	MI-BS 1	500	0.0816	0.0959	0.0861	0.0178	0.1484	-0.1005	0.2756
100	0.10	0.10	MI-BS 2	500	0.0819	0.0957	0.0859	0.0167	0.1480	-0.0932	0.2745
100	0.10	0.10	BS-MI	500	0.0021	0.1124	0.0041	-0.0759	0.0847	-0.2121	0.2133
100	0.10	0.10	MID-BS 1	500	0.0673	0.1011	0.0704	-0.0029	0.1377	-0.1242	0.2724
100	0.10	0.10	MID-BS 2	500	0.0672	0.1005	0.0699	-0.0020	0.1381	-0.1258	0.2732
100	0.10	0.10	BS-MID	500	0.0052	0.1122	0.0063	-0.0749	0.0864	-0.2122	0.2151
100	0.25	0.02	MI-BS 1	500	0.0129	0.0623	0.0160	-0.0262	0.0541	-0.1194	0.1311
100	0.25	0.02	MI-BS 2	500	0.0128	0.0623	0.0165	-0.0268	0.0531	-0.1196	0.1301
100	0.25	0.02	BS-MI	500	0.0053	0.0643	0.0078	-0.0357	0.0503	-0.1280	0.1249
100	0.25	0.02	MID-BS 1	500	0.0093	0.0627	0.0157	-0.0302	0.0517	-0.1247	0.1221
100	0.25	0.02	MID-BS 2	500	0.0094	0.0626	0.0143	-0.0313	0.0511	-0.1249	0.1249
100	0.25	0.02	BS-MID	500	0.0049	0.0640	0.0099	-0.0357	0.0496	-0.1293	0.1240
100	0.25	0.05	MI-BS 1	500	0.0210	0.0658	0.0276	-0.0201	0.0681	-0.1164	0.1267
100	0.25	0.05	MI-BS 2	500	0.0206	0.0657	0.0276	-0.0200	0.0680	-0.1211	0.1288
100	0.25	0.05	BS-MI	500	0.0012	0.0710	0.0071	-0.0448	0.0536	-0.1533	0.1148
100	0.25	0.05	MID-BS 1	500	0.0142	0.0669	0.0221	-0.0272	0.0625	-0.1251	0.1291
100	0.25	0.05	MID-BS 2	500	0.0142	0.0666	0.0211	-0.0277	0.0623	-0.1257	0.1254
100	0.25	0.05	BS-MID	500	0.0024	0.0700	0.0101	-0.0428	0.0515	-0.1444	0.1175
100	0.25	0.10	MI-BS 1	500	0.0415	0.0689	0.0429	-0.0020	0.0920	-0.1003	0.1607
100	0.25	0.10	MI-BS 2	500	0.0417	0.0692	0.0439	-0.0043	0.0900	-0.1007	0.1624
100	0.25	0.10	BS-MI	500	0.0022	0.0792	0.0065	-0.0515	0.0592	-0.1622	0.1448
100	0.25	0.10	MID-BS 1	500	0.0309	0.0710	0.0305	-0.0149	0.0792	-0.1123	0.1597
100	0.25	0.10	MID-BS 2	500	0.0306	0.0708	0.0312	-0.0168	0.0803	-0.1109	0.1580
100	0.25	0.10	BS-MID	500	0.0035	0.0779	0.0050	-0.0494	0.0573	-0.1563	0.1479
100	0.50	0.02	MI-BS 1	500	0.0052	0.0564	0.0062	-0.0269	0.0443	-0.1173	0.1031

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.50	0.02	MI-BS 2	500	0.0053	0.0561	0.0063	-0.0257	0.0452	-0.1121	0.1064
100	0.50	0.02	BS-MI	500	0.0014	0.0574	0.0017	-0.0321	0.0418	-0.1202	0.1069
100	0.50	0.02	MID-BS 1	500	0.0032	0.0562	0.0047	-0.0301	0.0418	-0.1172	0.1030
100	0.50	0.02	MID-BS 2	500	0.0031	0.0563	0.0036	-0.0301	0.0430	-0.1123	0.0987
100	0.50	0.02	BS-MID	500	0.0015	0.0568	0.0021	-0.0324	0.0425	-0.1171	0.1058
100	0.50	0.05	MI-BS 1	500	0.0093	0.0576	0.0107	-0.0254	0.0475	-0.1156	0.1110
100	0.50	0.05	MI-BS 2	500	0.0091	0.0578	0.0111	-0.0255	0.0469	-0.1176	0.1135
100	0.50	0.05	BS-MI	500	-0.0013	0.0605	0.0027	-0.0378	0.0405	-0.1363	0.1042
100	0.50	0.05	MID-BS 1	500	0.0039	0.0584	0.0058	-0.0309	0.0428	-0.1176	0.1127
100	0.50	0.05	MID-BS 2	500	0.0041	0.0585	0.0062	-0.0313	0.0430	-0.1193	0.1107
100	0.50	0.05	BS-MID	500	-0.0012	0.0596	-0.0001	-0.0372	0.0407	-0.1299	0.1084
100	0.50	0.10	MI-BS 1	500	0.0160	0.0575	0.0154	-0.0160	0.0531	-0.1003	0.1238
100	0.50	0.10	MI-BS 2	500	0.0159	0.0576	0.0168	-0.0178	0.0531	-0.1021	0.1246
100	0.50	0.10	BS-MI	500	-0.0061	0.0628	-0.0049	-0.0458	0.0361	-0.1311	0.1090
100	0.50	0.10	MID-BS 1	500	0.0069	0.0589	0.0098	-0.0289	0.0493	-0.1180	0.1170
100	0.50	0.10	MID-BS 2	500	0.0067	0.0591	0.0087	-0.0287	0.0482	-0.1147	0.1171
100	0.50	0.10	BS-MID	500	-0.0054	0.0613	-0.0019	-0.0410	0.0357	-0.1297	0.1079
500	0.10	0.02	MI-BS 1	500	0.0035	0.0409	0.0055	-0.0219	0.0317	-0.0785	0.0812
500	0.10	0.02	MI-BS 2	500	0.0036	0.0408	0.0061	-0.0228	0.0310	-0.0807	0.0802
500	0.10	0.02	BS-MI	500	-0.0016	0.0425	0.0017	-0.0303	0.0260	-0.0888	0.0780
500	0.10	0.02	MID-BS 1	500	0.0011	0.0415	0.0039	-0.0249	0.0285	-0.0846	0.0765
500	0.10	0.02	MID-BS 2	500	0.0010	0.0413	0.0038	-0.0253	0.0293	-0.0842	0.0782
500	0.10	0.02	BS-MID	500	-0.0025	0.0423	0.0002	-0.0275	0.0261	-0.0896	0.0756
500	0.10	0.05	MI-BS 1	500	0.0082	0.0452	0.0111	-0.0202	0.0397	-0.0827	0.0945
500	0.10	0.05	MI-BS 2	500	0.0084	0.0453	0.0108	-0.0206	0.0398	-0.0814	0.0929
500	0.10	0.05	BS-MI	500	-0.0065	0.0495	-0.0041	-0.0372	0.0262	-0.1086	0.0882
500	0.10	0.05	MID-BS 1	500	0.0034	0.0455	0.0055	-0.0239	0.0348	-0.0942	0.0861
500	0.10	0.05	MID-BS 2	500	0.0034	0.0454	0.0056	-0.0247	0.0350	-0.0952	0.0850
500	0.10	0.05	BS-MID	500	-0.0075	0.0487	-0.0058	-0.0370	0.0261	-0.1141	0.0779
500	0.10	0.10	MI-BS 1	500	0.0235	0.0493	0.0241	-0.0089	0.0584	-0.0811	0.1107
500	0.10	0.10	MI-BS 2	500	0.0233	0.0491	0.0237	-0.0083	0.0574	-0.0812	0.1136
500	0.10	0.10	BS-MI	500	-0.0080	0.0580	-0.0073	-0.0433	0.0347	-0.1354	0.0941
500	0.10	0.10	MID-BS 1	500	0.0157	0.0503	0.0167	-0.0170	0.0514	-0.0927	0.1068
500	0.10	0.10	MID-BS 2	500	0.0156	0.0503	0.0160	-0.0172	0.0514	-0.0920	0.1087
500	0.10	0.10	BS-MID	500	-0.0085	0.0572	-0.0074	-0.0473	0.0302	-0.1298	0.0942
500	0.25	0.02	MI-BS 1	500	0.0009	0.0271	0.0014	-0.0180	0.0196	-0.0527	0.0483
500	0.25	0.02	MI-BS 2	500	0.0010	0.0272	0.0014	-0.0170	0.0196	-0.0530	0.0493
500	0.25	0.02	BS-MI	500	-0.0011	0.0277	-0.0012	-0.0191	0.0190	-0.0551	0.0479
500	0.25	0.02	MID-BS 1	500	-0.0009	0.0273	-0.0007	-0.0197	0.0178	-0.0548	0.0477

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
500	0.25	0.02	MID-BS 2	500	-0.0009	0.0271	0.0001	-0.0193	0.0174	-0.0554	0.0475
500	0.25	0.02	BS-MID	500	-0.0022	0.0275	-0.0002	-0.0201	0.0170	-0.0565	0.0471
500	0.25	0.05	MI-BS 1	500	0.0018	0.0287	0.0034	-0.0172	0.0213	-0.0534	0.0536
500	0.25	0.05	MI-BS 2	500	0.0018	0.0289	0.0025	-0.0168	0.0220	-0.0531	0.0545
500	0.25	0.05	BS-MI	500	-0.0030	0.0302	-0.0018	-0.0221	0.0178	-0.0595	0.0516
500	0.25	0.05	MID-BS 1	500	-0.0013	0.0290	-0.0006	-0.0212	0.0195	-0.0562	0.0516
500	0.25	0.05	MID-BS 2	500	-0.0013	0.0291	-0.0010	-0.0213	0.0193	-0.0555	0.0518
500	0.25	0.05	BS-MID	500	-0.0044	0.0298	-0.0044	-0.0246	0.0164	-0.0602	0.0492
500	0.25	0.10	MI-BS 1	500	0.0057	0.0313	0.0093	-0.0182	0.0256	-0.0578	0.0623
500	0.25	0.10	MI-BS 2	500	0.0058	0.0313	0.0094	-0.0174	0.0260	-0.0584	0.0624
500	0.25	0.10	BS-MI	500	-0.0051	0.0344	-0.0021	-0.0311	0.0182	-0.0743	0.0588
500	0.25	0.10	MID-BS 1	500	0.0006	0.0319	0.0037	-0.0226	0.0221	-0.0638	0.0594
500	0.25	0.10	MID-BS 2	500	0.0005	0.0318	0.0038	-0.0232	0.0211	-0.0645	0.0588
500	0.25	0.10	BS-MID	500	-0.0068	0.0340	-0.0040	-0.0312	0.0153	-0.0763	0.0577
500	0.50	0.02	MI-BS 1	500	0.0007	0.0228	0.0020	-0.0140	0.0160	-0.0421	0.0452
500	0.50	0.02	MI-BS 2	500	0.0010	0.0226	0.0019	-0.0133	0.0158	-0.0425	0.0449
500	0.50	0.02	BS-MI	500	0.0000	0.0230	0.0008	-0.0148	0.0146	-0.0443	0.0456
500	0.50	0.02	MID-BS 1	500	-0.0003	0.0225	0.0004	-0.0159	0.0147	-0.0417	0.0421
500	0.50	0.02	MID-BS 2	500	-0.0004	0.0224	0.0008	-0.0158	0.0143	-0.0421	0.0429
500	0.50	0.02	BS-MID	500	-0.0009	0.0228	0.0006	-0.0165	0.0138	-0.0433	0.0439
500	0.50	0.05	MI-BS 1	500	0.0020	0.0235	0.0016	-0.0141	0.0184	-0.0455	0.0466
500	0.50	0.05	MI-BS 2	500	0.0019	0.0235	0.0023	-0.0144	0.0182	-0.0448	0.0448
500	0.50	0.05	BS-MI	500	-0.0001	0.0244	0.0003	-0.0166	0.0164	-0.0497	0.0459
500	0.50	0.05	MID-BS 1	500	-0.0009	0.0236	-0.0002	-0.0178	0.0158	-0.0483	0.0414
500	0.50	0.05	MID-BS 2	500	-0.0008	0.0234	0.0004	-0.0174	0.0152	-0.0480	0.0427
500	0.50	0.05	BS-MID	500	-0.0015	0.0238	-0.0005	-0.0180	0.0148	-0.0483	0.0416
500	0.50	0.10	MI-BS 1	500	0.0028	0.0252	0.0034	-0.0148	0.0199	-0.0457	0.0512
500	0.50	0.10	MI-BS 2	500	0.0030	0.0251	0.0034	-0.0148	0.0191	-0.0460	0.0517
500	0.50	0.10	BS-MI	500	-0.0025	0.0267	-0.0027	-0.0213	0.0151	-0.0534	0.0498
500	0.50	0.10	MID-BS 1	500	-0.0013	0.0253	-0.0004	-0.0181	0.0156	-0.0507	0.0471
500	0.50	0.10	MID-BS 2	500	-0.0014	0.0251	-0.0015	-0.0187	0.0150	-0.0496	0.0457
500	0.50	0.10	BS-MID	500	-0.0042	0.0258	-0.0039	-0.0227	0.0126	-0.0541	0.0439
1000	0.10	0.02	MI-BS 1	500	0.0029	0.0273	0.0039	-0.0162	0.0225	-0.0489	0.0559
1000	0.10	0.02	MI-BS 2	500	0.0030	0.0274	0.0038	-0.0155	0.0228	-0.0492	0.0563
1000	0.10	0.02	BS-MI	500	-0.0001	0.0283	0.0005	-0.0208	0.0204	-0.0540	0.0554
1000	0.10	0.02	MID-BS 1	500	0.0012	0.0271	0.0023	-0.0187	0.0210	-0.0502	0.0537
1000	0.10	0.02	MID-BS 2	500	0.0010	0.0272	0.0015	-0.0183	0.0208	-0.0503	0.0535
1000	0.10	0.02	BS-MID	500	-0.0011	0.0279	-0.0003	-0.0215	0.0194	-0.0547	0.0536
1000	0.10	0.05	MI-BS 1	500	0.0069	0.0280	0.0070	-0.0117	0.0261	-0.0553	0.0591

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.10	0.05	MI-BS 2	500	0.0070	0.0279	0.0061	-0.0121	0.0262	-0.0545	0.0583
1000	0.10	0.05	BS-MI	500	-0.0007	0.0300	-0.0013	-0.0213	0.0204	-0.0639	0.0541
1000	0.10	0.05	MID-BS 1	500	0.0030	0.0283	0.0030	-0.0166	0.0231	-0.0575	0.0559
1000	0.10	0.05	MID-BS 2	500	0.0032	0.0280	0.0030	-0.0164	0.0240	-0.0553	0.0552
1000	0.10	0.05	BS-MID	500	-0.0025	0.0297	-0.0030	-0.0231	0.0186	-0.0643	0.0513
1000	0.10	0.10	MI-BS 1	500	0.0125	0.0349	0.0135	-0.0098	0.0346	-0.0594	0.0808
1000	0.10	0.10	MI-BS 2	500	0.0124	0.0347	0.0132	-0.0098	0.0345	-0.0581	0.0807
1000	0.10	0.10	BS-MI	500	-0.0054	0.0402	-0.0053	-0.0305	0.0205	-0.0882	0.0728
1000	0.10	0.10	MID-BS 1	500	0.0071	0.0352	0.0077	-0.0148	0.0307	-0.0634	0.0776
1000	0.10	0.10	MID-BS 2	500	0.0070	0.0350	0.0080	-0.0159	0.0300	-0.0645	0.0764
1000	0.10	0.10	BS-MID	500	-0.0063	0.0393	-0.0059	-0.0306	0.0183	-0.0849	0.0736
1000	0.25	0.02	MI-BS 1	500	0.0007	0.0176	0.0011	-0.0111	0.0122	-0.0357	0.0334
1000	0.25	0.02	MI-BS 2	500	0.0008	0.0175	0.0011	-0.0104	0.0124	-0.0353	0.0335
1000	0.25	0.02	BS-MI	500	-0.0001	0.0179	-0.0002	-0.0112	0.0115	-0.0356	0.0329
1000	0.25	0.02	MID-BS 1	500	-0.0007	0.0178	-0.0003	-0.0122	0.0114	-0.0381	0.0312
1000	0.25	0.02	MID-BS 2	500	-0.0007	0.0176	0.0000	-0.0117	0.0111	-0.0362	0.0310
1000	0.25	0.02	BS-MID	500	-0.0011	0.0179	-0.0008	-0.0128	0.0108	-0.0367	0.0315
1000	0.25	0.05	MI-BS 1	500	0.0011	0.0189	0.0015	-0.0111	0.0139	-0.0368	0.0369
1000	0.25	0.05	MI-BS 2	500	0.0010	0.0187	0.0011	-0.0110	0.0135	-0.0368	0.0369
1000	0.25	0.05	BS-MI	500	-0.0014	0.0197	-0.0011	-0.0149	0.0113	-0.0415	0.0355
1000	0.25	0.05	MID-BS 1	500	-0.0017	0.0189	-0.0012	-0.0143	0.0110	-0.0389	0.0338
1000	0.25	0.05	MID-BS 2	500	-0.0016	0.0187	-0.0009	-0.0139	0.0112	-0.0403	0.0333
1000	0.25	0.05	BS-MID	500	-0.0031	0.0195	-0.0023	-0.0162	0.0100	-0.0428	0.0338
1000	0.25	0.10	MI-BS 1	500	0.0037	0.0217	0.0041	-0.0107	0.0180	-0.0407	0.0436
1000	0.25	0.10	MI-BS 2	500	0.0038	0.0216	0.0037	-0.0100	0.0176	-0.0398	0.0425
1000	0.25	0.10	BS-MI	500	-0.0022	0.0239	-0.0023	-0.0170	0.0144	-0.0528	0.0422
1000	0.25	0.10	MID-BS 1	500	-0.0004	0.0216	-0.0006	-0.0143	0.0145	-0.0427	0.0405
1000	0.25	0.10	MID-BS 2	500	-0.0002	0.0216	-0.0003	-0.0143	0.0145	-0.0433	0.0406
1000	0.25	0.10	BS-MID	500	-0.0042	0.0232	-0.0042	-0.0183	0.0116	-0.0508	0.0410
1000	0.50	0.02	MI-BS 1	500	0.0000	0.0157	-0.0003	-0.0103	0.0102	-0.0311	0.0288
1000	0.50	0.02	MI-BS 2	500	0.0001	0.0155	0.0003	-0.0103	0.0107	-0.0296	0.0290
1000	0.50	0.02	BS-MI	500	-0.0004	0.0159	-0.0001	-0.0111	0.0110	-0.0314	0.0293
1000	0.50	0.02	MID-BS 1	500	-0.0010	0.0155	-0.0012	-0.0118	0.0092	-0.0304	0.0282
1000	0.50	0.02	MID-BS 2	500	-0.0010	0.0155	-0.0010	-0.0122	0.0098	-0.0308	0.0275
1000	0.50	0.02	BS-MID	500	-0.0012	0.0157	-0.0013	-0.0121	0.0099	-0.0315	0.0282
1000	0.50	0.05	MI-BS 1	500	0.0001	0.0160	-0.0002	-0.0112	0.0118	-0.0312	0.0307
1000	0.50	0.05	MI-BS 2	500	0.0000	0.0160	-0.0001	-0.0109	0.0117	-0.0325	0.0302
1000	0.50	0.05	BS-MI	500	-0.0012	0.0163	-0.0011	-0.0124	0.0104	-0.0355	0.0298
1000	0.50	0.05	MID-BS 1	500	-0.0023	0.0160	-0.0024	-0.0133	0.0093	-0.0356	0.0259

Estadísticos descriptivos diferencia  $\Delta$  estadístico C (MAR)

<b>n</b>	<b>propY</b>	<b>propmiss</b>	<b>Estrategia</b>	<b>NSIM</b>	<b>Media</b>	<b>SD</b>	<b>Mediana</b>	<b>Q1</b>	<b>Q3</b>	<b>P2.5</b>	<b>P97.5</b>
1000	0.50	0.05	MID-BS 2	500	-0.0024	0.0160	-0.0027	-0.0127	0.0092	-0.0355	0.0260
1000	0.50	0.05	BS-MID	500	-0.0029	0.0161	-0.0026	-0.0135	0.0081	-0.0369	0.0259
1000	0.50	0.10	MI-BS 1	500	0.0014	0.0172	0.0013	-0.0106	0.0139	-0.0310	0.0326
1000	0.50	0.10	MI-BS 2	500	0.0014	0.0173	0.0015	-0.0110	0.0145	-0.0317	0.0326
1000	0.50	0.10	BS-MI	500	-0.0014	0.0185	-0.0009	-0.0148	0.0123	-0.0365	0.0326
1000	0.50	0.10	MID-BS 1	500	-0.0019	0.0171	-0.0011	-0.0136	0.0099	-0.0367	0.0307
1000	0.50	0.10	MID-BS 2	500	-0.0021	0.0171	-0.0011	-0.0138	0.0097	-0.0363	0.0292
1000	0.50	0.10	BS-MID	500	-0.0037	0.0178	-0.0034	-0.0166	0.0087	-0.0377	0.0302

**Anexo 15. Diferencia  $\Delta$  calibration slope por estrategias: estadísticos descriptivos bajo MCAR**

Estadísticos descriptivos diferencia  $\Delta$  calibration slope (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.10	0.02	MI-BS 1	500	0.0375	0.2830	0.0990	-0.0922	0.2255	-0.6590	0.4231
100	0.10	0.02	MI-BS 2	500	0.0385	0.2803	0.0893	-0.0809	0.2275	-0.6577	0.4130
100	0.10	0.02	BS-MI	500	0.0184	0.2874	0.0733	-0.1066	0.2091	-0.6783	0.4190
100	0.10	0.02	MID-BS 1	500	0.0343	0.2806	0.0946	-0.0877	0.2273	-0.6616	0.4234
100	0.10	0.02	MID-BS 2	500	0.0329	0.2788	0.0926	-0.0871	0.2186	-0.6685	0.4017
100	0.10	0.02	BS-MID	500	0.0228	0.2812	0.0857	-0.0940	0.2076	-0.6560	0.4111
100	0.10	0.05	MI-BS 1	500	0.0568	0.2907	0.1205	-0.0703	0.2479	-0.6080	0.4553
100	0.10	0.05	MI-BS 2	500	0.0567	0.2889	0.1176	-0.0655	0.2472	-0.6222	0.4471
100	0.10	0.05	BS-MI	500	0.0034	0.3047	0.0612	-0.1319	0.2166	-0.6910	0.4224
100	0.10	0.05	MID-BS 1	500	0.0467	0.2822	0.1081	-0.0719	0.2267	-0.6000	0.4434
100	0.10	0.05	MID-BS 2	500	0.0440	0.2827	0.1117	-0.0767	0.2254	-0.6035	0.4285
100	0.10	0.05	BS-MID	500	0.0169	0.2910	0.0807	-0.1213	0.2157	-0.6347	0.4091
100	0.10	0.10	MI-BS 1	500	0.1118	0.2653	0.1621	-0.0053	0.2857	-0.5186	0.4737
100	0.10	0.10	MI-BS 2	500	0.1121	0.2656	0.1575	0.0025	0.2864	-0.5498	0.4734
100	0.10	0.10	BS-MI	500	0.0108	0.2970	0.0601	-0.1325	0.1991	-0.6521	0.4234
100	0.10	0.10	MID-BS 1	500	0.0797	0.2672	0.1248	-0.0251	0.2542	-0.5495	0.4414
100	0.10	0.10	MID-BS 2	500	0.0804	0.2680	0.1316	-0.0304	0.2553	-0.5449	0.4276
100	0.10	0.10	BS-MID	500	0.0280	0.2850	0.0739	-0.0915	0.2152	-0.6187	0.4016
100	0.25	0.02	MI-BS 1	500	0.0261	0.2787	0.0812	-0.0798	0.2065	-0.6751	0.3949
100	0.25	0.02	MI-BS 2	500	0.0264	0.2777	0.0802	-0.0794	0.2050	-0.6884	0.3937
100	0.25	0.02	BS-MI	500	0.0083	0.2827	0.0663	-0.1061	0.1898	-0.6975	0.3843
100	0.25	0.02	MID-BS 1	500	0.0207	0.2811	0.0723	-0.0893	0.2052	-0.6895	0.3930
100	0.25	0.02	MID-BS 2	500	0.0221	0.2788	0.0751	-0.0827	0.2027	-0.6920	0.3917
100	0.25	0.02	BS-MID	500	0.0118	0.2823	0.0622	-0.1025	0.2001	-0.6907	0.3846
100	0.25	0.05	MI-BS 1	500	0.0513	0.2644	0.0993	-0.0666	0.2241	-0.6441	0.4165
100	0.25	0.05	MI-BS 2	500	0.0510	0.2646	0.1020	-0.0642	0.2223	-0.6440	0.4017
100	0.25	0.05	BS-MI	500	0.0068	0.2739	0.0624	-0.1118	0.1930	-0.6871	0.3802
100	0.25	0.05	MID-BS 1	500	0.0382	0.2713	0.0830	-0.0776	0.2133	-0.6452	0.4003
100	0.25	0.05	MID-BS 2	500	0.0371	0.2705	0.0798	-0.0759	0.2153	-0.6300	0.3971
100	0.25	0.05	BS-MID	500	0.0117	0.2749	0.0570	-0.1057	0.1983	-0.6499	0.3804
100	0.25	0.10	MI-BS 1	500	0.0717	0.2833	0.1339	-0.0316	0.2497	-0.7404	0.4531
100	0.25	0.10	MI-BS 2	500	0.0719	0.2818	0.1311	-0.0331	0.2464	-0.7250	0.4388
100	0.25	0.10	BS-MI	500	-0.0168	0.3028	0.0522	-0.1326	0.1716	-0.8696	0.3830
100	0.25	0.10	MID-BS 1	500	0.0508	0.2841	0.1092	-0.0580	0.2332	-0.7011	0.4232
100	0.25	0.10	MID-BS 2	500	0.0507	0.2837	0.1164	-0.0604	0.2259	-0.6923	0.4136
100	0.25	0.10	BS-MID	500	-0.0022	0.2932	0.0648	-0.1167	0.1816	-0.7663	0.4023
100	0.50	0.02	MI-BS 1	500	0.0109	0.2462	0.0455	-0.1054	0.1857	-0.6614	0.3557

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.50	0.02	MI-BS 2	500	0.0127	0.2456	0.0484	-0.1047	0.1826	-0.6561	0.3614
100	0.50	0.02	BS-MI	500	-0.0034	0.2494	0.0305	-0.1185	0.1724	-0.6852	0.3526
100	0.50	0.02	MID-BS 1	500	0.0080	0.2493	0.0415	-0.1095	0.1722	-0.6649	0.3709
100	0.50	0.02	MID-BS 2	500	0.0076	0.2484	0.0462	-0.1130	0.1741	-0.6588	0.3550
100	0.50	0.02	BS-MID	500	-0.0020	0.2502	0.0354	-0.1130	0.1654	-0.6738	0.3556
100	0.50	0.05	MI-BS 1	500	0.0336	0.2454	0.0609	-0.0740	0.2032	-0.6104	0.3935
100	0.50	0.05	MI-BS 2	500	0.0326	0.2444	0.0599	-0.0752	0.2001	-0.6211	0.3959
100	0.50	0.05	BS-MI	500	-0.0055	0.2552	0.0241	-0.1130	0.1776	-0.6622	0.3687
100	0.50	0.05	MID-BS 1	500	0.0205	0.2526	0.0530	-0.0861	0.1961	-0.6655	0.3843
100	0.50	0.05	MID-BS 2	500	0.0198	0.2520	0.0575	-0.0803	0.1923	-0.6560	0.3847
100	0.50	0.05	BS-MID	500	-0.0014	0.2577	0.0288	-0.1151	0.1794	-0.6901	0.3628
100	0.50	0.10	MI-BS 1	500	0.0661	0.2354	0.0995	-0.0540	0.2285	-0.5096	0.4161
100	0.50	0.10	MI-BS 2	500	0.0675	0.2356	0.1032	-0.0516	0.2285	-0.5049	0.4186
100	0.50	0.10	BS-MI	500	-0.0182	0.2569	0.0179	-0.1520	0.1661	-0.6622	0.3886
100	0.50	0.10	MID-BS 1	500	0.0417	0.2444	0.0761	-0.0844	0.2107	-0.5579	0.3997
100	0.50	0.10	MID-BS 2	500	0.0410	0.2441	0.0740	-0.0880	0.2122	-0.5658	0.3982
100	0.50	0.10	BS-MID	500	-0.0104	0.2569	0.0233	-0.1439	0.1641	-0.6380	0.3781
500	0.10	0.02	MI-BS 1	500	-0.0024	0.1977	0.0283	-0.0909	0.1323	-0.4444	0.2743
500	0.10	0.02	MI-BS 2	500	-0.0021	0.1977	0.0324	-0.0963	0.1315	-0.4449	0.2761
500	0.10	0.02	BS-MI	500	-0.0090	0.1995	0.0266	-0.0979	0.1228	-0.4502	0.2837
500	0.10	0.02	MID-BS 1	500	-0.0036	0.1977	0.0320	-0.0970	0.1267	-0.4472	0.2740
500	0.10	0.02	MID-BS 2	500	-0.0029	0.1965	0.0299	-0.0912	0.1288	-0.4375	0.2797
500	0.10	0.02	BS-MID	500	-0.0071	0.1974	0.0242	-0.1054	0.1259	-0.4493	0.2818
500	0.10	0.05	MI-BS 1	500	0.0001	0.1980	0.0383	-0.1124	0.1453	-0.4404	0.2948
500	0.10	0.05	MI-BS 2	500	-0.0001	0.1974	0.0357	-0.1130	0.1404	-0.4549	0.2977
500	0.10	0.05	BS-MI	500	-0.0177	0.2058	0.0157	-0.1341	0.1325	-0.5100	0.2967
500	0.10	0.05	MID-BS 1	500	-0.0036	0.2003	0.0290	-0.1228	0.1398	-0.4733	0.2929
500	0.10	0.05	MID-BS 2	500	-0.0040	0.1993	0.0297	-0.1126	0.1392	-0.4699	0.2977
500	0.10	0.05	BS-MID	500	-0.0148	0.2042	0.0155	-0.1320	0.1410	-0.4858	0.3032
500	0.10	0.10	MI-BS 1	500	0.0118	0.2115	0.0351	-0.0921	0.1572	-0.4895	0.3255
500	0.10	0.10	MI-BS 2	500	0.0113	0.2107	0.0385	-0.0927	0.1591	-0.5002	0.3211
500	0.10	0.10	BS-MI	500	-0.0301	0.2292	0.0075	-0.1337	0.1280	-0.5791	0.3238
500	0.10	0.10	MID-BS 1	500	0.0060	0.2100	0.0418	-0.1046	0.1495	-0.4853	0.3193
500	0.10	0.10	MID-BS 2	500	0.0056	0.2097	0.0373	-0.1022	0.1490	-0.4701	0.3264
500	0.10	0.10	BS-MID	500	-0.0178	0.2204	0.0176	-0.1346	0.1320	-0.5379	0.3262
500	0.25	0.02	MI-BS 1	500	0.0009	0.1310	0.0155	-0.0795	0.0959	-0.2968	0.2081
500	0.25	0.02	MI-BS 2	500	0.0004	0.1303	0.0126	-0.0784	0.0917	-0.3044	0.2007
500	0.25	0.02	BS-MI	500	-0.0046	0.1319	0.0136	-0.0861	0.0890	-0.3166	0.2005
500	0.25	0.02	MID-BS 1	500	0.0005	0.1310	0.0126	-0.0779	0.0924	-0.3062	0.2040

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
500	0.25	0.02	MID-BS 2	500	-0.0003	0.1303	0.0142	-0.0779	0.0904	-0.2962	0.1999
500	0.25	0.02	BS-MID	500	-0.0026	0.1304	0.0110	-0.0854	0.0892	-0.3071	0.2024
500	0.25	0.05	MI-BS 1	500	0.0096	0.1326	0.0309	-0.0685	0.1000	-0.2829	0.2376
500	0.25	0.05	MI-BS 2	500	0.0089	0.1328	0.0287	-0.0710	0.0982	-0.2871	0.2366
500	0.25	0.05	BS-MI	500	-0.0035	0.1380	0.0183	-0.0922	0.0918	-0.3033	0.2353
500	0.25	0.05	MID-BS 1	500	0.0082	0.1322	0.0239	-0.0790	0.0996	-0.2866	0.2373
500	0.25	0.05	MID-BS 2	500	0.0080	0.1316	0.0229	-0.0764	0.0964	-0.2808	0.2395
500	0.25	0.05	BS-MID	500	0.0011	0.1352	0.0231	-0.0827	0.0928	-0.2836	0.2378
500	0.25	0.10	MI-BS 1	500	0.0123	0.1426	0.0249	-0.0667	0.1100	-0.2957	0.2446
500	0.25	0.10	MI-BS 2	500	0.0121	0.1424	0.0236	-0.0622	0.1135	-0.2909	0.2420
500	0.25	0.10	BS-MI	500	-0.0115	0.1534	0.0001	-0.0953	0.0973	-0.3487	0.2515
500	0.25	0.10	MID-BS 1	500	0.0058	0.1438	0.0178	-0.0681	0.1050	-0.3270	0.2338
500	0.25	0.10	MID-BS 2	500	0.0063	0.1432	0.0176	-0.0654	0.1051	-0.3292	0.2304
500	0.25	0.10	BS-MID	500	-0.0074	0.1498	0.0089	-0.0864	0.0943	-0.3488	0.2296
500	0.50	0.02	MI-BS 1	500	0.0027	0.1131	0.0095	-0.0719	0.0849	-0.2322	0.2019
500	0.50	0.02	MI-BS 2	500	0.0028	0.1133	0.0134	-0.0738	0.0837	-0.2295	0.2009
500	0.50	0.02	BS-MI	500	-0.0011	0.1157	0.0104	-0.0792	0.0826	-0.2390	0.1965
500	0.50	0.02	MID-BS 1	500	0.0010	0.1125	0.0113	-0.0750	0.0798	-0.2349	0.2001
500	0.50	0.02	MID-BS 2	500	0.0008	0.1130	0.0110	-0.0755	0.0840	-0.2313	0.1966
500	0.50	0.02	BS-MID	500	-0.0010	0.1146	0.0109	-0.0797	0.0803	-0.2386	0.1987
500	0.50	0.05	MI-BS 1	500	0.0050	0.1226	0.0167	-0.0704	0.0964	-0.2496	0.2125
500	0.50	0.05	MI-BS 2	500	0.0054	0.1221	0.0150	-0.0725	0.0946	-0.2486	0.2100
500	0.50	0.05	BS-MI	500	-0.0051	0.1285	0.0062	-0.0884	0.0873	-0.2685	0.2075
500	0.50	0.05	MID-BS 1	500	0.0027	0.1226	0.0087	-0.0767	0.0962	-0.2499	0.2076
500	0.50	0.05	MID-BS 2	500	0.0031	0.1230	0.0097	-0.0799	0.0971	-0.2506	0.2018
500	0.50	0.05	BS-MID	500	-0.0028	0.1274	0.0059	-0.0855	0.0912	-0.2553	0.2092
500	0.50	0.10	MI-BS 1	500	0.0156	0.1272	0.0271	-0.0652	0.1139	-0.2588	0.2305
500	0.50	0.10	MI-BS 2	500	0.0151	0.1271	0.0272	-0.0645	0.1118	-0.2596	0.2266
500	0.50	0.10	BS-MI	500	-0.0054	0.1367	0.0029	-0.0940	0.0941	-0.3049	0.2253
500	0.50	0.10	MID-BS 1	500	0.0092	0.1267	0.0211	-0.0657	0.0982	-0.2722	0.2167
500	0.50	0.10	MID-BS 2	500	0.0090	0.1266	0.0212	-0.0720	0.1012	-0.2677	0.2191
500	0.50	0.10	BS-MID	500	-0.0029	0.1323	0.0049	-0.0803	0.0912	-0.2945	0.2166
1000	0.10	0.02	MI-BS 1	500	0.0091	0.1226	0.0136	-0.0711	0.0992	-0.2534	0.2096
1000	0.10	0.02	MI-BS 2	500	0.0090	0.1215	0.0114	-0.0684	0.1010	-0.2482	0.2069
1000	0.10	0.02	BS-MI	500	0.0054	0.1244	0.0114	-0.0720	0.0965	-0.2609	0.2088
1000	0.10	0.02	MID-BS 1	500	0.0086	0.1219	0.0117	-0.0623	0.0992	-0.2520	0.2100
1000	0.10	0.02	MID-BS 2	500	0.0079	0.1212	0.0098	-0.0688	0.1009	-0.2519	0.2077
1000	0.10	0.02	BS-MID	500	0.0058	0.1227	0.0148	-0.0709	0.0997	-0.2534	0.2085
1000	0.10	0.05	MI-BS 1	500	0.0112	0.1247	0.0244	-0.0693	0.1011	-0.2560	0.2078

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.10	0.05	MI-BS 2	500	0.0104	0.1238	0.0223	-0.0637	0.0982	-0.2635	0.2092
1000	0.10	0.05	BS-MI	500	0.0006	0.1301	0.0094	-0.0833	0.0951	-0.2914	0.2195
1000	0.10	0.05	MID-BS 1	500	0.0089	0.1243	0.0228	-0.0671	0.0979	-0.2561	0.2055
1000	0.10	0.05	MID-BS 2	500	0.0097	0.1238	0.0240	-0.0671	0.1003	-0.2539	0.2060
1000	0.10	0.05	BS-MID	500	0.0042	0.1271	0.0163	-0.0745	0.0906	-0.2655	0.2054
1000	0.10	0.10	MI-BS 1	500	0.0125	0.1359	0.0284	-0.0725	0.1078	-0.2893	0.2435
1000	0.10	0.10	MI-BS 2	500	0.0127	0.1354	0.0299	-0.0750	0.1055	-0.2905	0.2340
1000	0.10	0.10	BS-MI	500	-0.0095	0.1443	0.0065	-0.1051	0.0898	-0.3438	0.2423
1000	0.10	0.10	MID-BS 1	500	0.0105	0.1351	0.0241	-0.0756	0.1082	-0.2932	0.2356
1000	0.10	0.10	MID-BS 2	500	0.0103	0.1350	0.0251	-0.0777	0.1057	-0.2948	0.2330
1000	0.10	0.10	BS-MID	500	-0.0017	0.1409	0.0174	-0.0917	0.0909	-0.3294	0.2419
1000	0.25	0.02	MI-BS 1	500	0.0030	0.0892	0.0102	-0.0524	0.0635	-0.1760	0.1581
1000	0.25	0.02	MI-BS 2	500	0.0036	0.0885	0.0116	-0.0537	0.0607	-0.1786	0.1615
1000	0.25	0.02	BS-MI	500	0.0019	0.0899	0.0083	-0.0569	0.0593	-0.1837	0.1638
1000	0.25	0.02	MID-BS 1	500	0.0033	0.0895	0.0108	-0.0555	0.0613	-0.1770	0.1600
1000	0.25	0.02	MID-BS 2	500	0.0026	0.0888	0.0074	-0.0547	0.0620	-0.1734	0.1596
1000	0.25	0.02	BS-MID	500	0.0017	0.0900	0.0062	-0.0566	0.0604	-0.1843	0.1587
1000	0.25	0.05	MI-BS 1	500	0.0086	0.0931	0.0151	-0.0513	0.0684	-0.1822	0.1817
1000	0.25	0.05	MI-BS 2	500	0.0085	0.0925	0.0106	-0.0468	0.0713	-0.1777	0.1766
1000	0.25	0.05	BS-MI	500	0.0025	0.0954	0.0072	-0.0572	0.0651	-0.1864	0.1791
1000	0.25	0.05	MID-BS 1	500	0.0069	0.0929	0.0094	-0.0526	0.0666	-0.1826	0.1745
1000	0.25	0.05	MID-BS 2	500	0.0074	0.0927	0.0098	-0.0522	0.0673	-0.1872	0.1756
1000	0.25	0.05	BS-MID	500	0.0036	0.0945	0.0054	-0.0526	0.0655	-0.1881	0.1772
1000	0.25	0.10	MI-BS 1	500	0.0059	0.0974	0.0110	-0.0557	0.0726	-0.2026	0.1861
1000	0.25	0.10	MI-BS 2	500	0.0061	0.0980	0.0097	-0.0554	0.0718	-0.2119	0.1884
1000	0.25	0.10	BS-MI	500	-0.0077	0.1055	-0.0043	-0.0775	0.0655	-0.2346	0.1883
1000	0.25	0.10	MID-BS 1	500	0.0034	0.0972	0.0065	-0.0570	0.0711	-0.2104	0.1858
1000	0.25	0.10	MID-BS 2	500	0.0040	0.0965	0.0076	-0.0559	0.0702	-0.2129	0.1846
1000	0.25	0.10	BS-MID	500	-0.0036	0.1016	0.0005	-0.0685	0.0669	-0.2312	0.1804
1000	0.50	0.02	MI-BS 1	500	0.0038	0.0799	0.0063	-0.0486	0.0605	-0.1621	0.1534
1000	0.50	0.02	MI-BS 2	500	0.0037	0.0793	0.0062	-0.0467	0.0593	-0.1679	0.1477
1000	0.50	0.02	BS-MI	500	0.0018	0.0805	0.0047	-0.0478	0.0586	-0.1816	0.1486
1000	0.50	0.02	MID-BS 1	500	0.0031	0.0793	0.0060	-0.0469	0.0602	-0.1728	0.1401
1000	0.50	0.02	MID-BS 2	500	0.0028	0.0792	0.0053	-0.0462	0.0593	-0.1747	0.1459
1000	0.50	0.02	BS-MID	500	0.0016	0.0799	0.0045	-0.0501	0.0602	-0.1803	0.1464
1000	0.50	0.05	MI-BS 1	500	0.0072	0.0819	0.0139	-0.0453	0.0637	-0.1755	0.1499
1000	0.50	0.05	MI-BS 2	500	0.0076	0.0820	0.0144	-0.0449	0.0650	-0.1715	0.1532
1000	0.50	0.05	BS-MI	500	0.0036	0.0853	0.0102	-0.0480	0.0647	-0.1802	0.1523
1000	0.50	0.05	MID-BS 1	500	0.0062	0.0825	0.0104	-0.0458	0.0617	-0.1761	0.1523

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MCAR)

<b>n</b>	<b>propY</b>	<b>propmiss</b>	<b>Estrategia</b>	<b>NSIM</b>	<b>Media</b>	<b>SD</b>	<b>Mediana</b>	<b>Q1</b>	<b>Q3</b>	<b>P2.5</b>	<b>P97.5</b>
1000	0.50	0.05	MID-BS 2	500	0.0060	0.0824	0.0118	-0.0442	0.0620	-0.1747	0.1505
1000	0.50	0.05	BS-MID	500	0.0040	0.0844	0.0088	-0.0460	0.0605	-0.1842	0.1511
1000	0.50	0.10	MI-BS 1	500	0.0097	0.0868	0.0123	-0.0478	0.0690	-0.1628	0.1595
1000	0.50	0.10	MI-BS 2	500	0.0099	0.0870	0.0141	-0.0456	0.0696	-0.1682	0.1577
1000	0.50	0.10	BS-MI	500	-0.0007	0.0930	0.0013	-0.0618	0.0634	-0.1819	0.1705
1000	0.50	0.10	MID-BS 1	500	0.0069	0.0869	0.0106	-0.0521	0.0646	-0.1656	0.1610
1000	0.50	0.10	MID-BS 2	500	0.0069	0.0863	0.0118	-0.0488	0.0626	-0.1649	0.1610
1000	0.50	0.10	BS-MID	500	0.0006	0.0894	0.0067	-0.0553	0.0613	-0.1832	0.1637

**Anexo 16. Diferencia  $\Delta$  calibration slope por estrategias: estadísticos descriptivos bajo MAR**

Estadísticos descriptivos diferencia  $\Delta$  calibration slope (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.10	0.02	MI-BS 1	500	0.0501	0.3200	0.1220	-0.0709	0.2553	-0.7808	0.4350
100	0.10	0.02	MI-BS 2	500	0.0501	0.3188	0.1248	-0.0746	0.2543	-0.7740	0.4539
100	0.10	0.02	BS-MI	500	-0.0026	0.3298	0.0671	-0.1334	0.2070	-0.8308	0.4112
100	0.10	0.02	MID-BS 1	500	0.0415	0.3196	0.1089	-0.0768	0.2422	-0.7471	0.4500
100	0.10	0.02	MID-BS 2	500	0.0402	0.3190	0.1124	-0.0748	0.2434	-0.7550	0.4264
100	0.10	0.02	BS-MID	500	0.0050	0.3278	0.0829	-0.1269	0.2139	-0.7918	0.4116
100	0.10	0.05	MI-BS 1	500	0.0943	0.2942	0.1427	-0.0665	0.2916	-0.5912	0.5034
100	0.10	0.05	MI-BS 2	500	0.0930	0.2917	0.1457	-0.0587	0.2921	-0.5864	0.4966
100	0.10	0.05	BS-MI	500	-0.0148	0.3407	0.0464	-0.1815	0.1995	-0.7858	0.4584
100	0.10	0.05	MID-BS 1	500	0.0693	0.3021	0.1342	-0.0710	0.2646	-0.6872	0.4709
100	0.10	0.05	MID-BS 2	500	0.0691	0.3016	0.1366	-0.0749	0.2698	-0.6694	0.4656
100	0.10	0.05	BS-MID	500	-0.0063	0.3345	0.0574	-0.1638	0.2095	-0.8038	0.4414
100	0.10	0.10	MI-BS 1	500	0.1673	0.2804	0.2226	0.0396	0.3514	-0.5254	0.5361
100	0.10	0.10	MI-BS 2	500	0.1670	0.2757	0.2192	0.0493	0.3471	-0.5172	0.5320
100	0.10	0.10	BS-MI	500	-0.0270	0.3258	0.0352	-0.1801	0.2075	-0.7855	0.4172
100	0.10	0.10	MID-BS 1	500	0.1267	0.2687	0.1874	-0.0047	0.2999	-0.5098	0.5179
100	0.10	0.10	MID-BS 2	500	0.1254	0.2658	0.1781	-0.0002	0.2957	-0.5185	0.4923
100	0.10	0.10	BS-MID	499	-0.0077	0.3154	0.0489	-0.1772	0.2012	-0.7353	0.4022
100	0.25	0.02	MI-BS 1	500	0.0393	0.2562	0.0828	-0.0732	0.2068	-0.5975	0.4116
100	0.25	0.02	MI-BS 2	500	0.0395	0.2559	0.0820	-0.0782	0.2061	-0.6036	0.4116
100	0.25	0.02	BS-MI	500	0.0125	0.2638	0.0456	-0.1258	0.1914	-0.6219	0.3937
100	0.25	0.02	MID-BS 1	500	0.0300	0.2582	0.0736	-0.0852	0.2026	-0.6200	0.3979
100	0.25	0.02	MID-BS 2	500	0.0305	0.2578	0.0698	-0.0887	0.2026	-0.6319	0.4027
100	0.25	0.02	BS-MID	500	0.0156	0.2637	0.0511	-0.1083	0.1967	-0.6368	0.4139
100	0.25	0.05	MI-BS 1	500	0.0595	0.2900	0.1279	-0.0415	0.2503	-0.7196	0.4044
100	0.25	0.05	MI-BS 2	500	0.0576	0.2896	0.1215	-0.0452	0.2440	-0.7377	0.4078
100	0.25	0.05	BS-MI	500	-0.0110	0.3092	0.0595	-0.1343	0.1939	-0.8404	0.3659
100	0.25	0.05	MID-BS 1	500	0.0408	0.2893	0.0968	-0.0585	0.2346	-0.6809	0.4118
100	0.25	0.05	MID-BS 2	500	0.0409	0.2885	0.0980	-0.0550	0.2313	-0.6722	0.4019
100	0.25	0.05	BS-MID	500	0.0011	0.3011	0.0669	-0.1156	0.2029	-0.7308	0.3776
100	0.25	0.10	MI-BS 1	500	0.1249	0.2684	0.1668	-0.0025	0.3111	-0.5295	0.4888
100	0.25	0.10	MI-BS 2	500	0.1259	0.2694	0.1728	0.0028	0.3122	-0.5221	0.4963
100	0.25	0.10	BS-MI	500	-0.0072	0.3002	0.0352	-0.1593	0.2107	-0.7202	0.4126
100	0.25	0.10	MID-BS 1	500	0.0957	0.2763	0.1478	-0.0398	0.2892	-0.5712	0.4691
100	0.25	0.10	MID-BS 2	500	0.0953	0.2747	0.1507	-0.0377	0.2879	-0.5714	0.4603
100	0.25	0.10	BS-MID	500	0.0083	0.2962	0.0582	-0.1448	0.2184	-0.7150	0.4118
100	0.50	0.02	MI-BS 1	500	0.0093	0.2546	0.0419	-0.0969	0.1847	-0.6502	0.3721

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.50	0.02	MI-BS 2	500	0.0104	0.2528	0.0461	-0.1057	0.1778	-0.6555	0.3826
100	0.50	0.02	BS-MI	500	-0.0064	0.2588	0.0371	-0.1194	0.1644	-0.6552	0.3790
100	0.50	0.02	MID-BS 1	500	0.0071	0.2493	0.0406	-0.1042	0.1780	-0.6333	0.3805
100	0.50	0.02	MID-BS 2	500	0.0065	0.2488	0.0463	-0.1063	0.1723	-0.6322	0.3805
100	0.50	0.02	BS-MID	500	-0.0007	0.2507	0.0388	-0.1149	0.1677	-0.6331	0.3724
100	0.50	0.05	MI-BS 1	500	0.0246	0.2682	0.0723	-0.0839	0.2014	-0.6352	0.4090
100	0.50	0.05	MI-BS 2	500	0.0234	0.2699	0.0863	-0.0867	0.1991	-0.6505	0.4149
100	0.50	0.05	BS-MI	500	-0.0170	0.2807	0.0401	-0.1320	0.1653	-0.7142	0.3792
100	0.50	0.05	MID-BS 1	500	0.0110	0.2662	0.0559	-0.1017	0.1919	-0.6254	0.3959
100	0.50	0.05	MID-BS 2	500	0.0122	0.2666	0.0584	-0.0955	0.1915	-0.6285	0.4045
100	0.50	0.05	BS-MID	500	-0.0081	0.2720	0.0425	-0.1248	0.1745	-0.6602	0.3855
100	0.50	0.10	MI-BS 1	500	0.0499	0.2668	0.0920	-0.0603	0.2276	-0.5739	0.4313
100	0.50	0.10	MI-BS 2	500	0.0493	0.2665	0.0938	-0.0680	0.2227	-0.5794	0.4335
100	0.50	0.10	BS-MI	500	-0.0350	0.2864	0.0106	-0.1507	0.1511	-0.7037	0.3650
100	0.50	0.10	MID-BS 1	500	0.0251	0.2696	0.0645	-0.0995	0.2132	-0.6122	0.4016
100	0.50	0.10	MID-BS 2	500	0.0240	0.2695	0.0641	-0.1005	0.2140	-0.5974	0.3989
100	0.50	0.10	BS-MID	500	-0.0220	0.2783	0.0179	-0.1359	0.1675	-0.6839	0.3586
500	0.10	0.02	MI-BS 1	500	0.0054	0.1924	0.0411	-0.0936	0.1336	-0.4488	0.2966
500	0.10	0.02	MI-BS 2	500	0.0055	0.1927	0.0406	-0.0998	0.1318	-0.4537	0.2967
500	0.10	0.02	BS-MI	500	-0.0152	0.2007	0.0154	-0.1186	0.1184	-0.4860	0.2891
500	0.10	0.02	MID-BS 1	500	0.0006	0.1948	0.0346	-0.0992	0.1321	-0.4955	0.2964
500	0.10	0.02	MID-BS 2	500	0.0005	0.1940	0.0381	-0.0973	0.1279	-0.4982	0.3011
500	0.10	0.02	BS-MID	500	-0.0130	0.1989	0.0223	-0.1187	0.1197	-0.5027	0.2975
500	0.10	0.05	MI-BS 1	500	0.0127	0.2307	0.0535	-0.0977	0.1634	-0.4731	0.3412
500	0.10	0.05	MI-BS 2	500	0.0130	0.2313	0.0566	-0.0868	0.1677	-0.4654	0.3319
500	0.10	0.05	BS-MI	500	-0.0441	0.2517	-0.0028	-0.1664	0.1197	-0.5630	0.3188
500	0.10	0.05	MID-BS 1	500	0.0032	0.2309	0.0423	-0.1069	0.1571	-0.5350	0.3255
500	0.10	0.05	MID-BS 2	500	0.0029	0.2299	0.0417	-0.1034	0.1557	-0.5313	0.3235
500	0.10	0.05	BS-MID	500	-0.0377	0.2462	0.0050	-0.1565	0.1246	-0.6119	0.3142
500	0.10	0.10	MI-BS 1	500	0.0675	0.2287	0.1052	-0.0428	0.2256	-0.5127	0.4063
500	0.10	0.10	MI-BS 2	500	0.0671	0.2279	0.1071	-0.0394	0.2232	-0.5068	0.4048
500	0.10	0.10	BS-MI	500	-0.0500	0.2672	-0.0171	-0.1716	0.1372	-0.7006	0.3549
500	0.10	0.10	MID-BS 1	500	0.0536	0.2299	0.0937	-0.0658	0.2147	-0.5365	0.3900
500	0.10	0.10	MID-BS 2	500	0.0529	0.2301	0.0943	-0.0720	0.2103	-0.5273	0.3952
500	0.10	0.10	BS-MID	500	-0.0340	0.2598	-0.0002	-0.1682	0.1472	-0.6890	0.3501
500	0.25	0.02	MI-BS 1	500	0.0040	0.1334	0.0196	-0.0728	0.0964	-0.2977	0.2181
500	0.25	0.02	MI-BS 2	500	0.0046	0.1332	0.0192	-0.0731	0.1000	-0.2960	0.2189
500	0.25	0.02	BS-MI	500	-0.0040	0.1361	0.0125	-0.0840	0.0881	-0.2957	0.2188
500	0.25	0.02	MID-BS 1	500	0.0024	0.1336	0.0164	-0.0731	0.0942	-0.2980	0.2168

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
500	0.25	0.02	MID-BS 2	500	0.0021	0.1330	0.0204	-0.0720	0.0966	-0.2940	0.2133
500	0.25	0.02	BS-MID	500	-0.0030	0.1354	0.0128	-0.0876	0.0894	-0.2978	0.2223
500	0.25	0.05	MI-BS 1	500	0.0053	0.1410	0.0218	-0.0741	0.1060	-0.3136	0.2380
500	0.25	0.05	MI-BS 2	500	0.0056	0.1417	0.0156	-0.0755	0.1085	-0.3127	0.2422
500	0.25	0.05	BS-MI	500	-0.0149	0.1490	-0.0008	-0.1042	0.0953	-0.3492	0.2286
500	0.25	0.05	MID-BS 1	500	0.0011	0.1410	0.0093	-0.0854	0.1063	-0.2922	0.2473
500	0.25	0.05	MID-BS 2	500	0.0010	0.1415	0.0100	-0.0830	0.0986	-0.3065	0.2368
500	0.25	0.05	BS-MID	500	-0.0118	0.1460	0.0016	-0.1025	0.0971	-0.3274	0.2359
500	0.25	0.10	MI-BS 1	500	0.0199	0.1561	0.0463	-0.0805	0.1254	-0.3265	0.2629
500	0.25	0.10	MI-BS 2	500	0.0202	0.1561	0.0463	-0.0830	0.1256	-0.3455	0.2663
500	0.25	0.10	BS-MI	500	-0.0242	0.1720	-0.0010	-0.1286	0.0982	-0.4192	0.2460
500	0.25	0.10	MID-BS 1	500	0.0101	0.1582	0.0378	-0.0894	0.1160	-0.3533	0.2583
500	0.25	0.10	MID-BS 2	500	0.0097	0.1582	0.0342	-0.0947	0.1178	-0.3510	0.2555
500	0.25	0.10	BS-MID	500	-0.0191	0.1696	0.0059	-0.1220	0.0944	-0.4139	0.2504
500	0.50	0.02	MI-BS 1	500	0.0008	0.1137	0.0118	-0.0781	0.0874	-0.2325	0.2042
500	0.50	0.02	MI-BS 2	500	0.0020	0.1126	0.0093	-0.0753	0.0888	-0.2314	0.2016
500	0.50	0.02	BS-MI	500	-0.0021	0.1143	0.0058	-0.0831	0.0815	-0.2310	0.2042
500	0.50	0.02	MID-BS 1	500	0.0017	0.1126	0.0080	-0.0799	0.0876	-0.2314	0.2022
500	0.50	0.02	MID-BS 2	500	0.0012	0.1119	0.0093	-0.0815	0.0860	-0.2236	0.1997
500	0.50	0.02	BS-MID	500	-0.0008	0.1136	0.0081	-0.0806	0.0838	-0.2331	0.2042
500	0.50	0.05	MI-BS 1	500	0.0061	0.1171	0.0124	-0.0720	0.0958	-0.2450	0.2086
500	0.50	0.05	MI-BS 2	500	0.0052	0.1171	0.0128	-0.0750	0.0950	-0.2596	0.2035
500	0.50	0.05	BS-MI	500	-0.0035	0.1221	0.0055	-0.0822	0.0868	-0.2831	0.2042
500	0.50	0.05	MID-BS 1	500	0.0027	0.1169	0.0145	-0.0803	0.0896	-0.2509	0.2006
500	0.50	0.05	MID-BS 2	500	0.0033	0.1159	0.0145	-0.0782	0.0918	-0.2428	0.2077
500	0.50	0.05	BS-MID	500	0.0005	0.1175	0.0141	-0.0788	0.0896	-0.2565	0.2041
500	0.50	0.10	MI-BS 1	500	0.0093	0.1278	0.0203	-0.0769	0.0970	-0.2628	0.2386
500	0.50	0.10	MI-BS 2	500	0.0101	0.1275	0.0220	-0.0748	0.0969	-0.2651	0.2377
500	0.50	0.10	BS-MI	500	-0.0137	0.1362	-0.0022	-0.1034	0.0801	-0.2973	0.2413
500	0.50	0.10	MID-BS 1	500	0.0030	0.1263	0.0144	-0.0782	0.0873	-0.2679	0.2331
500	0.50	0.10	MID-BS 2	500	0.0029	0.1257	0.0106	-0.0766	0.0890	-0.2702	0.2270
500	0.50	0.10	BS-MID	500	-0.0090	0.1300	0.0041	-0.0976	0.0789	-0.2823	0.2208
1000	0.10	0.02	MI-BS 1	500	0.0106	0.1281	0.0183	-0.0695	0.1006	-0.2723	0.2281
1000	0.10	0.02	MI-BS 2	500	0.0108	0.1283	0.0199	-0.0685	0.1019	-0.2832	0.2301
1000	0.10	0.02	BS-MI	500	-0.0029	0.1337	0.0041	-0.0845	0.0973	-0.2970	0.2243
1000	0.10	0.02	MID-BS 1	500	0.0096	0.1264	0.0189	-0.0705	0.0989	-0.2612	0.2265
1000	0.10	0.02	MID-BS 2	500	0.0087	0.1265	0.0203	-0.0708	0.0998	-0.2719	0.2278
1000	0.10	0.02	BS-MID	500	0.0001	0.1309	0.0075	-0.0850	0.0961	-0.2951	0.2185
1000	0.10	0.05	MI-BS 1	500	0.0276	0.1308	0.0322	-0.0542	0.1256	-0.2722	0.2370

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.10	0.05	MI-BS 2	500	0.0278	0.1301	0.0335	-0.0543	0.1240	-0.2771	0.2333
1000	0.10	0.05	BS-MI	500	-0.0021	0.1428	0.0084	-0.0855	0.1001	-0.3294	0.2313
1000	0.10	0.05	MID-BS 1	500	0.0225	0.1298	0.0342	-0.0602	0.1156	-0.2635	0.2295
1000	0.10	0.05	MID-BS 2	500	0.0235	0.1287	0.0312	-0.0558	0.1181	-0.2621	0.2279
1000	0.10	0.05	BS-MID	500	0.0021	0.1374	0.0107	-0.0814	0.0980	-0.2995	0.2351
1000	0.10	0.10	MI-BS 1	500	0.0424	0.1653	0.0663	-0.0588	0.1545	-0.3605	0.3083
1000	0.10	0.10	MI-BS 2	500	0.0423	0.1644	0.0623	-0.0559	0.1553	-0.3470	0.3091
1000	0.10	0.10	BS-MI	500	-0.0279	0.1894	-0.0035	-0.1375	0.1092	-0.4724	0.2917
1000	0.10	0.10	MID-BS 1	500	0.0352	0.1657	0.0519	-0.0538	0.1461	-0.3454	0.2992
1000	0.10	0.10	MID-BS 2	500	0.0348	0.1643	0.0540	-0.0526	0.1432	-0.3497	0.2995
1000	0.10	0.10	BS-MID	500	-0.0165	0.1834	0.0042	-0.1117	0.1132	-0.4591	0.2826
1000	0.25	0.02	MI-BS 1	500	0.0032	0.0904	0.0086	-0.0558	0.0650	-0.1918	0.1672
1000	0.25	0.02	MI-BS 2	500	0.0035	0.0902	0.0057	-0.0568	0.0669	-0.1924	0.1667
1000	0.25	0.02	BS-MI	500	0.0003	0.0923	0.0014	-0.0600	0.0626	-0.1935	0.1632
1000	0.25	0.02	MID-BS 1	500	0.0018	0.0919	0.0051	-0.0589	0.0633	-0.1968	0.1656
1000	0.25	0.02	MID-BS 2	500	0.0018	0.0910	0.0044	-0.0542	0.0645	-0.1969	0.1680
1000	0.25	0.02	BS-MID	500	0.0006	0.0925	0.0038	-0.0586	0.0613	-0.1926	0.1654
1000	0.25	0.05	MI-BS 1	500	0.0059	0.0981	0.0134	-0.0552	0.0711	-0.2121	0.1805
1000	0.25	0.05	MI-BS 2	500	0.0057	0.0971	0.0139	-0.0536	0.0702	-0.2072	0.1803
1000	0.25	0.05	BS-MI	500	-0.0047	0.1028	0.0029	-0.0737	0.0616	-0.2270	0.1778
1000	0.25	0.05	MID-BS 1	500	0.0024	0.0977	0.0124	-0.0567	0.0665	-0.2041	0.1791
1000	0.25	0.05	MID-BS 2	500	0.0031	0.0967	0.0099	-0.0572	0.0670	-0.1976	0.1748
1000	0.25	0.05	BS-MID	500	-0.0033	0.1005	0.0044	-0.0690	0.0622	-0.2125	0.1748
1000	0.25	0.10	MI-BS 1	500	0.0171	0.1098	0.0235	-0.0455	0.0940	-0.2349	0.2033
1000	0.25	0.10	MI-BS 2	500	0.0173	0.1093	0.0246	-0.0467	0.0912	-0.2314	0.1989
1000	0.25	0.10	BS-MI	500	-0.0071	0.1205	0.0001	-0.0778	0.0774	-0.2624	0.1900
1000	0.25	0.10	MID-BS 1	500	0.0117	0.1091	0.0172	-0.0561	0.0884	-0.2244	0.1967
1000	0.25	0.10	MID-BS 2	500	0.0126	0.1088	0.0170	-0.0535	0.0865	-0.2226	0.2001
1000	0.25	0.10	BS-MID	500	-0.0033	0.1161	0.0026	-0.0713	0.0795	-0.2543	0.1989
1000	0.50	0.02	MI-BS 1	500	0.0019	0.0808	0.0056	-0.0527	0.0601	-0.1817	0.1458
1000	0.50	0.02	MI-BS 2	500	0.0025	0.0801	0.0067	-0.0517	0.0594	-0.1769	0.1438
1000	0.50	0.02	BS-MI	500	0.0001	0.0820	0.0048	-0.0565	0.0588	-0.1862	0.1427
1000	0.50	0.02	MID-BS 1	500	0.0026	0.0802	0.0081	-0.0505	0.0606	-0.1747	0.1474
1000	0.50	0.02	MID-BS 2	500	0.0027	0.0800	0.0076	-0.0525	0.0609	-0.1781	0.1415
1000	0.50	0.02	BS-MID	500	0.0020	0.0809	0.0076	-0.0516	0.0593	-0.1817	0.1458
1000	0.50	0.05	MI-BS 1	500	0.0023	0.0833	0.0079	-0.0560	0.0631	-0.1679	0.1487
1000	0.50	0.05	MI-BS 2	500	0.0017	0.0831	0.0064	-0.0547	0.0612	-0.1659	0.1498
1000	0.50	0.05	BS-MI	500	-0.0036	0.0847	0.0011	-0.0566	0.0590	-0.1741	0.1416
1000	0.50	0.05	MID-BS 1	500	0.0005	0.0831	0.0036	-0.0547	0.0613	-0.1686	0.1467

Estadísticos descriptivos diferencia  $\Delta$  *calibration slope* (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.50	0.05	MID-BS 2	500	0.0002	0.0829	0.0047	-0.0554	0.0609	-0.1786	0.1456
1000	0.50	0.05	BS-MID	500	-0.0023	0.0835	0.0031	-0.0571	0.0608	-0.1796	0.1413
1000	0.50	0.10	MI-BS 1	500	0.0081	0.0902	0.0149	-0.0543	0.0737	-0.1805	0.1679
1000	0.50	0.10	MI-BS 2	500	0.0081	0.0909	0.0157	-0.0520	0.0740	-0.1813	0.1650
1000	0.50	0.10	BS-MI	500	-0.0042	0.0972	0.0055	-0.0691	0.0644	-0.2020	0.1605
1000	0.50	0.10	MID-BS 1	500	0.0055	0.0896	0.0107	-0.0558	0.0699	-0.1870	0.1592
1000	0.50	0.10	MID-BS 2	500	0.0049	0.0896	0.0124	-0.0550	0.0687	-0.1871	0.1608
1000	0.50	0.10	BS-MID	500	-0.0019	0.0930	0.0079	-0.0599	0.0611	-0.1942	0.1627

### Anexo 17. Diferencia $\Delta$ CITL por estrategias: estadísticos descriptivos bajo MCAR

Estadísticos descriptivos diferencia $\Delta$ CITL (MCAR)											
n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.10	0.02	MI-BS 1	187	0.0677	0.1811	0.0565	-0.0591	0.1665	-0.2564	0.4634
100	0.10	0.02	MI-BS 2	70	0.0717	0.1518	0.0649	-0.0338	0.1681	-0.1474	0.3784
100	0.10	0.02	BS-MI	176	0.0526	0.1822	0.0438	-0.0882	0.1579	-0.2410	0.4253
100	0.10	0.02	MID-BS 1	168	0.0625	0.1668	0.0574	-0.0505	0.1445	-0.2063	0.3940
100	0.10	0.02	MID-BS 2	50	0.0366	0.1294	0.0288	-0.0427	0.1178	-0.1837	0.3462
100	0.10	0.02	BS-MID	172	0.0550	0.1765	0.0362	-0.0678	0.1526	-0.2357	0.4621
100	0.10	0.05	MI-BS 1	110	0.0705	0.1902	0.0535	-0.0668	0.1902	-0.2081	0.4652
100	0.10	0.05	MI-BS 2	59	0.0973	0.1696	0.0793	-0.0055	0.1922	-0.2254	0.4482
100	0.10	0.05	BS-MI	119	0.1045	0.2009	0.1014	-0.0308	0.2256	-0.2618	0.5835
100	0.10	0.05	MID-BS 1	117	0.0723	0.1743	0.0651	-0.0508	0.1666	-0.1983	0.4147
100	0.10	0.05	MID-BS 2	33	0.1063	0.1443	0.0526	-0.0022	0.1964	-0.1358	0.3858
100	0.10	0.05	BS-MID	109	0.0893	0.1807	0.0673	-0.0316	0.2087	-0.1970	0.4468
100	0.10	0.10	MI-BS 1	85	0.1557	0.2362	0.1174	-0.0174	0.2851	-0.1755	0.6102
100	0.10	0.10	MI-BS 2	60	0.1268	0.1481	0.1061	0.0331	0.2421	-0.1622	0.3938
100	0.10	0.10	BS-MI	72	0.1401	0.2161	0.1467	-0.0012	0.2471	-0.2733	0.6003
100	0.10	0.10	MID-BS 1	54	0.0233	0.1796	0.0399	-0.0927	0.1131	-0.3003	0.3660
100	0.10	0.10	MID-BS 2	18	0.1041	0.1489	0.0956	0.0054	0.1965	-0.1285	0.3577
100	0.10	0.10	BS-MID	45	0.1129	0.2022	0.1037	0.0216	0.2204	-0.2969	0.4719
100	0.25	0.02	MI-BS 1	499	0.0245	0.1476	0.0262	-0.0741	0.1183	-0.2634	0.3007
100	0.25	0.02	MI-BS 2	494	0.0238	0.1442	0.0315	-0.0734	0.1109	-0.2673	0.3048
100	0.25	0.02	BS-MI	495	0.0212	0.1507	0.0225	-0.0777	0.1190	-0.2640	0.2927
100	0.25	0.02	MID-BS 1	498	0.0225	0.1492	0.0275	-0.0795	0.1128	-0.2692	0.3311
100	0.25	0.02	MID-BS 2	495	0.0228	0.1441	0.0261	-0.0704	0.1109	-0.2642	0.3024
100	0.25	0.02	BS-MID	494	0.0211	0.1473	0.0203	-0.0745	0.1182	-0.2520	0.2950
100	0.25	0.05	MI-BS 1	495	0.0214	0.1638	0.0148	-0.0798	0.1268	-0.3438	0.3357
100	0.25	0.05	MI-BS 2	491	0.0237	0.1605	0.0246	-0.0778	0.1277	-0.3339	0.3158
100	0.25	0.05	BS-MI	492	0.0180	0.1748	0.0248	-0.0929	0.1308	-0.3407	0.3390
100	0.25	0.05	MID-BS 1	493	0.0170	0.1591	0.0250	-0.0796	0.1159	-0.3130	0.3205
100	0.25	0.05	MID-BS 2	487	0.0211	0.1739	0.0220	-0.0800	0.1131	-0.3241	0.3077
100	0.25	0.05	BS-MID	491	0.0117	0.1629	0.0188	-0.0896	0.1152	-0.3323	0.3131
100	0.25	0.10	MI-BS 1	488	0.0279	0.1970	0.0366	-0.1035	0.1450	-0.3598	0.4241
100	0.25	0.10	MI-BS 2	480	0.0273	0.1970	0.0329	-0.0966	0.1389	-0.3516	0.4090
100	0.25	0.10	BS-MI	482	0.0064	0.2225	0.0032	-0.1169	0.1377	-0.4730	0.4642
100	0.25	0.10	MID-BS 1	472	0.0185	0.1937	0.0276	-0.1103	0.1292	-0.3445	0.4240
100	0.25	0.10	MID-BS 2	462	0.0228	0.1878	0.0310	-0.0998	0.1286	-0.3345	0.4202
100	0.25	0.10	BS-MID	474	0.0073	0.1969	0.0120	-0.1159	0.1206	-0.3638	0.4052
100	0.50	0.02	MI-BS 1	500	0.0112	0.1301	0.0123	-0.0742	0.0820	-0.2595	0.2994

Estadísticos descriptivos diferencia  $\Delta$  CITL (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.50	0.02	MI-BS 2	500	0.0135	0.1286	0.0125	-0.0683	0.0886	-0.2362	0.2995
100	0.50	0.02	BS-MI	500	0.0145	0.1321	0.0125	-0.0656	0.0950	-0.2572	0.2781
100	0.50	0.02	MID-BS 1	500	0.0145	0.1289	0.0210	-0.0732	0.0959	-0.2330	0.2927
100	0.50	0.02	MID-BS 2	500	0.0132	0.1291	0.0135	-0.0695	0.0904	-0.2569	0.2927
100	0.50	0.02	BS-MID	500	0.0144	0.1283	0.0158	-0.0709	0.0942	-0.2421	0.2998
100	0.50	0.05	MI-BS 1	500	0.0163	0.1451	0.0159	-0.0724	0.1038	-0.2849	0.3154
100	0.50	0.05	MI-BS 2	500	0.0137	0.1424	0.0105	-0.0768	0.1036	-0.2732	0.3056
100	0.50	0.05	BS-MI	500	0.0153	0.1535	0.0098	-0.0801	0.1152	-0.2881	0.3139
100	0.50	0.05	MID-BS 1	500	0.0123	0.1443	0.0075	-0.0774	0.0972	-0.2601	0.3105
100	0.50	0.05	MID-BS 2	500	0.0128	0.1432	0.0036	-0.0815	0.1058	-0.2785	0.3109
100	0.50	0.05	BS-MID	500	0.0146	0.1446	0.0064	-0.0811	0.1169	-0.2765	0.3043
100	0.50	0.10	MI-BS 1	500	0.0177	0.1693	0.0154	-0.0895	0.1178	-0.3093	0.3384
100	0.50	0.10	MI-BS 2	499	0.0162	0.1675	0.0139	-0.0857	0.1134	-0.3023	0.3439
100	0.50	0.10	BS-MI	500	0.0129	0.1900	0.0099	-0.1034	0.1191	-0.3842	0.4150
100	0.50	0.10	MID-BS 1	499	0.0203	0.1669	0.0147	-0.0868	0.1219	-0.2843	0.3532
100	0.50	0.10	MID-BS 2	499	0.0210	0.1704	0.0204	-0.0833	0.1178	-0.2980	0.3507
100	0.50	0.10	BS-MID	500	0.0167	0.1716	0.0152	-0.0852	0.1231	-0.3155	0.3516
500	0.10	0.02	MI-BS 1	500	-0.0078	0.0635	-0.0110	-0.0470	0.0346	-0.1361	0.1153
500	0.10	0.02	MI-BS 2	500	-0.0087	0.0630	-0.0106	-0.0479	0.0310	-0.1345	0.1211
500	0.10	0.02	BS-MI	500	-0.0095	0.0656	-0.0073	-0.0536	0.0308	-0.1365	0.1211
500	0.10	0.02	MID-BS 1	500	-0.0109	0.0640	-0.0120	-0.0496	0.0284	-0.1400	0.1161
500	0.10	0.02	MID-BS 2	500	-0.0083	0.0629	-0.0112	-0.0464	0.0296	-0.1307	0.1200
500	0.10	0.02	BS-MID	500	-0.0085	0.0634	-0.0065	-0.0500	0.0331	-0.1410	0.1271
500	0.10	0.05	MI-BS 1	500	-0.0068	0.0743	-0.0054	-0.0508	0.0383	-0.1687	0.1271
500	0.10	0.05	MI-BS 2	500	-0.0071	0.0735	-0.0080	-0.0530	0.0387	-0.1550	0.1313
500	0.10	0.05	BS-MI	500	-0.0074	0.0822	-0.0064	-0.0582	0.0436	-0.1860	0.1568
500	0.10	0.05	MID-BS 1	500	-0.0102	0.0735	-0.0117	-0.0555	0.0352	-0.1594	0.1310
500	0.10	0.05	MID-BS 2	500	-0.0105	0.0713	-0.0116	-0.0549	0.0358	-0.1611	0.1273
500	0.10	0.05	BS-MID	500	-0.0104	0.0745	-0.0128	-0.0540	0.0397	-0.1680	0.1332
500	0.10	0.10	MI-BS 1	500	-0.0063	0.0879	-0.0001	-0.0613	0.0504	-0.1823	0.1642
500	0.10	0.10	MI-BS 2	500	-0.0057	0.0857	-0.0016	-0.0583	0.0477	-0.1769	0.1649
500	0.10	0.10	BS-MI	500	-0.0059	0.1028	-0.0034	-0.0718	0.0624	-0.2094	0.1957
500	0.10	0.10	MID-BS 1	500	-0.0116	0.0815	-0.0086	-0.0625	0.0415	-0.1929	0.1475
500	0.10	0.10	MID-BS 2	500	-0.0120	0.0814	-0.0072	-0.0644	0.0382	-0.1821	0.1481
500	0.10	0.10	BS-MID	500	-0.0116	0.0857	-0.0077	-0.0623	0.0440	-0.1991	0.1590
500	0.25	0.02	MI-BS 1	500	0.0063	0.0516	0.0053	-0.0269	0.0404	-0.0879	0.1061
500	0.25	0.02	MI-BS 2	500	0.0061	0.0505	0.0045	-0.0242	0.0382	-0.0897	0.1062
500	0.25	0.02	BS-MI	500	0.0063	0.0532	0.0053	-0.0268	0.0421	-0.0951	0.1051
500	0.25	0.02	MID-BS 1	500	0.0041	0.0514	0.0040	-0.0300	0.0365	-0.0912	0.1073

Estadísticos descriptivos diferencia  $\Delta$  CITL (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
500	0.25	0.02	MID-BS 2	500	0.0059	0.0505	0.0060	-0.0287	0.0369	-0.0932	0.1080
500	0.25	0.02	BS-MID	500	0.0056	0.0521	0.0078	-0.0281	0.0395	-0.0956	0.1110
500	0.25	0.05	MI-BS 1	500	0.0042	0.0565	0.0078	-0.0347	0.0384	-0.1010	0.1130
500	0.25	0.05	MI-BS 2	500	0.0042	0.0558	0.0052	-0.0346	0.0401	-0.1056	0.1148
500	0.25	0.05	BS-MI	500	0.0040	0.0615	0.0041	-0.0383	0.0438	-0.1078	0.1226
500	0.25	0.05	MID-BS 1	500	0.0030	0.0568	0.0046	-0.0338	0.0383	-0.1095	0.1147
500	0.25	0.05	MID-BS 2	500	0.0035	0.0555	0.0039	-0.0311	0.0400	-0.1063	0.1113
500	0.25	0.05	BS-MID	500	0.0030	0.0580	0.0045	-0.0357	0.0424	-0.1189	0.1177
500	0.25	0.10	MI-BS 1	500	0.0098	0.0676	0.0080	-0.0304	0.0512	-0.1241	0.1500
500	0.25	0.10	MI-BS 2	500	0.0094	0.0671	0.0085	-0.0332	0.0530	-0.1185	0.1500
500	0.25	0.10	BS-MI	500	0.0105	0.0765	0.0105	-0.0400	0.0605	-0.1313	0.1691
500	0.25	0.10	MID-BS 1	500	0.0058	0.0658	0.0026	-0.0364	0.0502	-0.1168	0.1401
500	0.25	0.10	MID-BS 2	500	0.0062	0.0639	0.0049	-0.0363	0.0496	-0.1112	0.1343
500	0.25	0.10	BS-MID	500	0.0059	0.0669	0.0054	-0.0365	0.0471	-0.1253	0.1349
500	0.50	0.02	MI-BS 1	500	-0.0026	0.0510	-0.0038	-0.0382	0.0291	-0.1020	0.1028
500	0.50	0.02	MI-BS 2	500	-0.0017	0.0498	0.0002	-0.0343	0.0290	-0.0993	0.1033
500	0.50	0.02	BS-MI	500	-0.0009	0.0512	-0.0025	-0.0340	0.0328	-0.0961	0.1055
500	0.50	0.02	MID-BS 1	500	-0.0019	0.0504	-0.0002	-0.0367	0.0297	-0.1047	0.1048
500	0.50	0.02	MID-BS 2	500	-0.0021	0.0497	-0.0005	-0.0345	0.0288	-0.0985	0.1053
500	0.50	0.02	BS-MID	500	-0.0014	0.0493	-0.0011	-0.0322	0.0302	-0.0969	0.1011
500	0.50	0.05	MI-BS 1	500	-0.0033	0.0556	-0.0052	-0.0379	0.0345	-0.1151	0.1044
500	0.50	0.05	MI-BS 2	500	-0.0029	0.0556	-0.0058	-0.0380	0.0345	-0.1101	0.1077
500	0.50	0.05	BS-MI	500	-0.0036	0.0595	-0.0032	-0.0436	0.0348	-0.1171	0.1173
500	0.50	0.05	MID-BS 1	500	-0.0021	0.0567	-0.0039	-0.0401	0.0345	-0.1076	0.1163
500	0.50	0.05	MID-BS 2	500	-0.0028	0.0544	-0.0026	-0.0371	0.0321	-0.1095	0.1049
500	0.50	0.05	BS-MID	500	-0.0022	0.0549	-0.0028	-0.0364	0.0329	-0.1113	0.1086
500	0.50	0.10	MI-BS 1	500	-0.0017	0.0632	-0.0053	-0.0428	0.0393	-0.1219	0.1337
500	0.50	0.10	MI-BS 2	500	-0.0017	0.0625	-0.0064	-0.0437	0.0383	-0.1209	0.1295
500	0.50	0.10	BS-MI	500	-0.0006	0.0695	-0.0043	-0.0460	0.0437	-0.1324	0.1336
500	0.50	0.10	MID-BS 1	500	-0.0031	0.0623	-0.0073	-0.0423	0.0374	-0.1264	0.1284
500	0.50	0.10	MID-BS 2	500	-0.0025	0.0609	-0.0032	-0.0421	0.0358	-0.1172	0.1142
500	0.50	0.10	BS-MID	500	-0.0023	0.0628	-0.0026	-0.0442	0.0373	-0.1136	0.1254
1000	0.10	0.02	MI-BS 1	500	0.0029	0.0438	0.0021	-0.0265	0.0310	-0.0788	0.0924
1000	0.10	0.02	MI-BS 2	500	0.0033	0.0423	0.0024	-0.0258	0.0312	-0.0747	0.0872
1000	0.10	0.02	BS-MI	500	0.0048	0.0458	0.0060	-0.0276	0.0342	-0.0790	0.1009
1000	0.10	0.02	MID-BS 1	500	0.0026	0.0438	0.0010	-0.0268	0.0319	-0.0811	0.0882
1000	0.10	0.02	MID-BS 2	500	0.0032	0.0420	0.0028	-0.0267	0.0290	-0.0736	0.0880
1000	0.10	0.02	BS-MID	500	0.0041	0.0439	0.0031	-0.0268	0.0311	-0.0760	0.0926
1000	0.10	0.05	MI-BS 1	500	0.0022	0.0516	0.0032	-0.0281	0.0338	-0.1089	0.1055

Estadísticos descriptivos diferencia  $\Delta$  CITL (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.10	0.05	MI-BS 2	500	0.0026	0.0509	0.0044	-0.0296	0.0352	-0.1037	0.1045
1000	0.10	0.05	BS-MI	500	0.0023	0.0559	0.0012	-0.0326	0.0366	-0.1079	0.1125
1000	0.10	0.05	MID-BS 1	500	0.0020	0.0501	0.0022	-0.0289	0.0321	-0.0959	0.1033
1000	0.10	0.05	MID-BS 2	500	0.0019	0.0492	0.0041	-0.0275	0.0312	-0.1062	0.0940
1000	0.10	0.05	BS-MID	500	0.0014	0.0511	0.0041	-0.0311	0.0317	-0.0993	0.0993
1000	0.10	0.10	MI-BS 1	500	0.0036	0.0565	0.0025	-0.0345	0.0430	-0.1033	0.1107
1000	0.10	0.10	MI-BS 2	500	0.0041	0.0562	-0.0005	-0.0344	0.0449	-0.1020	0.1130
1000	0.10	0.10	BS-MI	500	0.0051	0.0628	0.0012	-0.0361	0.0493	-0.1101	0.1353
1000	0.10	0.10	MID-BS 1	500	-0.0008	0.0569	-0.0039	-0.0402	0.0396	-0.1100	0.1035
1000	0.10	0.10	MID-BS 2	500	-0.0009	0.0547	-0.0019	-0.0356	0.0380	-0.1087	0.1047
1000	0.10	0.10	BS-MID	500	-0.0006	0.0550	-0.0010	-0.0365	0.0355	-0.1058	0.1073
1000	0.25	0.02	MI-BS 1	500	0.0033	0.0370	0.0028	-0.0210	0.0271	-0.0618	0.0790
1000	0.25	0.02	MI-BS 2	500	0.0036	0.0365	0.0030	-0.0220	0.0258	-0.0639	0.0763
1000	0.25	0.02	BS-MI	500	0.0041	0.0379	0.0033	-0.0222	0.0277	-0.0671	0.0827
1000	0.25	0.02	MID-BS 1	500	0.0034	0.0368	0.0020	-0.0212	0.0251	-0.0596	0.0770
1000	0.25	0.02	MID-BS 2	500	0.0033	0.0360	0.0024	-0.0194	0.0250	-0.0632	0.0767
1000	0.25	0.02	BS-MID	500	0.0037	0.0366	0.0019	-0.0185	0.0257	-0.0658	0.0804
1000	0.25	0.05	MI-BS 1	500	0.0049	0.0410	0.0026	-0.0256	0.0318	-0.0671	0.0884
1000	0.25	0.05	MI-BS 2	500	0.0048	0.0404	0.0021	-0.0220	0.0306	-0.0694	0.0838
1000	0.25	0.05	BS-MI	500	0.0054	0.0449	0.0036	-0.0251	0.0352	-0.0779	0.0998
1000	0.25	0.05	MID-BS 1	500	0.0036	0.0401	0.0021	-0.0240	0.0301	-0.0718	0.0864
1000	0.25	0.05	MID-BS 2	500	0.0038	0.0390	0.0020	-0.0236	0.0286	-0.0738	0.0845
1000	0.25	0.05	BS-MID	500	0.0033	0.0410	0.0029	-0.0248	0.0304	-0.0738	0.0876
1000	0.25	0.10	MI-BS 1	500	0.0058	0.0483	0.0073	-0.0282	0.0358	-0.0884	0.1035
1000	0.25	0.10	MI-BS 2	500	0.0053	0.0477	0.0081	-0.0277	0.0355	-0.0855	0.1009
1000	0.25	0.10	BS-MI	500	0.0047	0.0542	0.0066	-0.0286	0.0406	-0.1025	0.1152
1000	0.25	0.10	MID-BS 1	500	0.0040	0.0472	0.0066	-0.0288	0.0354	-0.0843	0.0975
1000	0.25	0.10	MID-BS 2	500	0.0043	0.0462	0.0057	-0.0285	0.0341	-0.0859	0.1001
1000	0.25	0.10	BS-MID	500	0.0041	0.0477	0.0056	-0.0284	0.0323	-0.0878	0.1032
1000	0.50	0.02	MI-BS 1	500	0.0013	0.0351	0.0013	-0.0194	0.0250	-0.0692	0.0680
1000	0.50	0.02	MI-BS 2	500	0.0012	0.0349	0.0021	-0.0208	0.0239	-0.0675	0.0686
1000	0.50	0.02	BS-MI	500	0.0015	0.0367	0.0032	-0.0211	0.0247	-0.0703	0.0711
1000	0.50	0.02	MID-BS 1	500	0.0012	0.0351	0.0019	-0.0217	0.0234	-0.0690	0.0699
1000	0.50	0.02	MID-BS 2	500	0.0010	0.0345	0.0025	-0.0215	0.0238	-0.0681	0.0664
1000	0.50	0.02	BS-MID	500	0.0011	0.0355	0.0012	-0.0212	0.0241	-0.0736	0.0698
1000	0.50	0.05	MI-BS 1	500	0.0018	0.0386	0.0016	-0.0221	0.0278	-0.0733	0.0745
1000	0.50	0.05	MI-BS 2	500	0.0017	0.0381	0.0014	-0.0221	0.0277	-0.0727	0.0742
1000	0.50	0.05	BS-MI	500	0.0013	0.0405	0.0018	-0.0254	0.0277	-0.0769	0.0806
1000	0.50	0.05	MID-BS 1	500	0.0017	0.0378	0.0009	-0.0240	0.0255	-0.0707	0.0739

Estadísticos descriptivos diferencia  $\Delta$  CITL (MCAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.50	0.05	MID-BS 2	500	0.0019	0.0376	0.0010	-0.0235	0.0266	-0.0670	0.0745
1000	0.50	0.05	BS-MID	500	0.0016	0.0384	0.0008	-0.0227	0.0276	-0.0714	0.0792
1000	0.50	0.10	MI-BS 1	500	0.0011	0.0424	0.0010	-0.0279	0.0298	-0.0812	0.0865
1000	0.50	0.10	MI-BS 2	500	0.0008	0.0421	0.0002	-0.0273	0.0290	-0.0806	0.0792
1000	0.50	0.10	BS-MI	500	0.0000	0.0472	-0.0002	-0.0315	0.0301	-0.0930	0.0930
1000	0.50	0.10	MID-BS 1	500	0.0015	0.0411	0.0001	-0.0261	0.0298	-0.0784	0.0792
1000	0.50	0.10	MID-BS 2	500	0.0015	0.0408	0.0009	-0.0261	0.0298	-0.0802	0.0859
1000	0.50	0.10	BS-MID	500	0.0012	0.0423	0.0012	-0.0283	0.0304	-0.0825	0.0873

**Anexo 18. Diferencia  $\Delta$  CITL por estrategias: estadísticos descriptivos bajo MAR**

Estadísticos descriptivos diferencia $\Delta$ CITL (MAR)											
n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.10	0.02	MI-BS 1	128	0.0203	0.1845	0.0172	-0.1008	0.1447	-0.3468	0.3554
100	0.10	0.02	MI-BS 2	63	0.0520	0.1652	0.0320	-0.0798	0.1430	-0.2171	0.3804
100	0.10	0.02	BS-MI	129	0.0471	0.1851	0.0304	-0.0885	0.1625	-0.2630	0.4633
100	0.10	0.02	MID-BS 1	112	0.0335	0.1751	0.0413	-0.0799	0.1667	-0.3405	0.3513
100	0.10	0.02	MID-BS 2	44	0.0337	0.1555	0.0141	-0.0952	0.1339	-0.2190	0.3093
100	0.10	0.02	BS-MID	126	0.0423	0.1676	0.0245	-0.0681	0.1421	-0.2605	0.4025
100	0.10	0.05	MI-BS 1	67	0.0578	0.1516	0.0556	-0.0431	0.1659	-0.2100	0.3608
100	0.10	0.05	MI-BS 2	43	0.0481	0.1312	0.0725	-0.0397	0.1370	-0.2361	0.2916
100	0.10	0.05	BS-MI	56	0.0089	0.1738	0.0110	-0.0836	0.1093	-0.3421	0.3156
100	0.10	0.05	MID-BS 1	56	0.0211	0.1880	-0.0049	-0.0825	0.1363	-0.2583	0.3541
100	0.10	0.05	MID-BS 2	22	0.0482	0.1134	0.0646	-0.0168	0.1362	-0.1633	0.1994
100	0.10	0.05	BS-MID	43	0.0148	0.1723	0.0034	-0.0970	0.1057	-0.2860	0.3503
100	0.10	0.10	MI-BS 1	43	0.1062	0.2384	0.0792	-0.0362	0.2584	-0.2598	0.5826
100	0.10	0.10	MI-BS 2	33	0.1059	0.2166	0.0663	-0.0288	0.2043	-0.2052	0.5833
100	0.10	0.10	BS-MI	17	0.0321	0.2565	-0.0127	-0.1331	0.1976	-0.3390	0.5386
100	0.10	0.10	MID-BS 1	13	0.0057	0.2323	-0.0468	-0.0952	0.0430	-0.2140	0.5242
100	0.10	0.10	MID-BS 2	7	-0.0223	0.1800	-0.0470	-0.0921	0.0596	-0.2802	0.2406
100	0.10	0.10	BS-MID	4	0.0039	0.2404	0.0608	-0.0770	0.1418	-0.3057	0.2169
100	0.25	0.02	MI-BS 1	494	0.0137	0.1486	0.0144	-0.0852	0.1039	-0.2830	0.3207
100	0.25	0.02	MI-BS 2	497	0.0126	0.1494	0.0090	-0.0831	0.1020	-0.3000	0.3100
100	0.25	0.02	BS-MI	497	0.0186	0.1561	0.0271	-0.0853	0.1208	-0.2888	0.3128
100	0.25	0.02	MID-BS 1	495	0.0111	0.1511	0.0117	-0.0876	0.1070	-0.2864	0.3058
100	0.25	0.02	MID-BS 2	492	0.0128	0.1472	0.0061	-0.0777	0.1038	-0.2969	0.3200
100	0.25	0.02	BS-MID	497	0.0212	0.1522	0.0255	-0.0767	0.1123	-0.2987	0.3288
100	0.25	0.05	MI-BS 1	496	0.0138	0.1957	0.0028	-0.1018	0.1187	-0.3158	0.3806
100	0.25	0.05	MI-BS 2	490	0.0099	0.1726	-0.0019	-0.0981	0.1089	-0.3215	0.3774
100	0.25	0.05	BS-MI	488	0.0270	0.1863	0.0169	-0.0892	0.1352	-0.3247	0.4244
100	0.25	0.05	MID-BS 1	487	0.0041	0.1755	-0.0115	-0.1103	0.1150	-0.3422	0.3825
100	0.25	0.05	MID-BS 2	490	0.0080	0.1680	0.0002	-0.1031	0.1139	-0.3034	0.3732
100	0.25	0.05	BS-MID	484	0.0314	0.2126	0.0135	-0.0771	0.1278	-0.2935	0.4141
100	0.25	0.10	MI-BS 1	482	-0.0038	0.2101	-0.0075	-0.1457	0.1365	-0.3712	0.3910
100	0.25	0.10	MI-BS 2	476	-0.0054	0.2065	-0.0161	-0.1479	0.1216	-0.3611	0.4169
100	0.25	0.10	BS-MI	446	0.0361	0.2345	0.0266	-0.1252	0.1795	-0.3619	0.5126
100	0.25	0.10	MID-BS 1	451	-0.0184	0.2083	-0.0272	-0.1696	0.1116	-0.3747	0.4097
100	0.25	0.10	MID-BS 2	449	-0.0124	0.2073	-0.0241	-0.1562	0.1243	-0.3770	0.4048
100	0.25	0.10	BS-MID	416	0.0340	0.2172	0.0218	-0.1151	0.1683	-0.3595	0.5027
100	0.50	0.02	MI-BS 1	500	0.0106	0.1358	0.0101	-0.0828	0.1010	-0.2569	0.2881

Estadísticos descriptivos diferencia  $\Delta$  CITL (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
100	0.50	0.02	MI-BS 2	500	0.0111	0.1348	0.0112	-0.0806	0.0933	-0.2523	0.2871
100	0.50	0.02	BS-MI	500	0.0134	0.1422	0.0131	-0.0733	0.1068	-0.2915	0.3007
100	0.50	0.02	MID-BS 1	500	0.0139	0.1361	0.0149	-0.0783	0.0961	-0.2563	0.2910
100	0.50	0.02	MID-BS 2	500	0.0123	0.1337	0.0142	-0.0755	0.0919	-0.2479	0.2908
100	0.50	0.02	BS-MID	500	0.0153	0.1380	0.0134	-0.0721	0.1082	-0.2806	0.2975
100	0.50	0.05	MI-BS 1	500	0.0062	0.1523	0.0063	-0.0903	0.1004	-0.2937	0.3490
100	0.50	0.05	MI-BS 2	499	0.0074	0.1517	0.0103	-0.0875	0.0958	-0.2807	0.3424
100	0.50	0.05	BS-MI	499	0.0191	0.1639	0.0142	-0.0832	0.1194	-0.2832	0.3685
100	0.50	0.05	MID-BS 1	500	0.0086	0.1495	0.0098	-0.0889	0.1050	-0.2786	0.3508
100	0.50	0.05	MID-BS 2	500	0.0103	0.1491	0.0163	-0.0834	0.0998	-0.2789	0.3419
100	0.50	0.05	BS-MID	499	0.0224	0.1546	0.0239	-0.0769	0.1141	-0.2760	0.3528
100	0.50	0.10	MI-BS 1	500	-0.0082	0.1670	-0.0005	-0.1062	0.0998	-0.3474	0.3259
100	0.50	0.10	MI-BS 2	500	-0.0095	0.1654	-0.0025	-0.1107	0.0923	-0.3379	0.3171
100	0.50	0.10	BS-MI	499	0.0136	0.1828	0.0123	-0.1066	0.1291	-0.3507	0.3685
100	0.50	0.10	MID-BS 1	499	-0.0046	0.1647	-0.0012	-0.1098	0.0948	-0.3412	0.3193
100	0.50	0.10	MID-BS 2	498	-0.0070	0.1668	-0.0022	-0.1056	0.0817	-0.3253	0.3270
100	0.50	0.10	BS-MID	499	0.0185	0.1709	0.0224	-0.0926	0.1211	-0.3177	0.3592
500	0.10	0.02	MI-BS 1	500	-0.0095	0.0694	-0.0075	-0.0542	0.0386	-0.1448	0.1239
500	0.10	0.02	MI-BS 2	500	-0.0094	0.0697	-0.0098	-0.0527	0.0354	-0.1591	0.1215
500	0.10	0.02	BS-MI	500	-0.0051	0.0749	-0.0027	-0.0528	0.0405	-0.1582	0.1386
500	0.10	0.02	MID-BS 1	500	-0.0104	0.0708	-0.0139	-0.0537	0.0356	-0.1555	0.1259
500	0.10	0.02	MID-BS 2	500	-0.0111	0.0673	-0.0137	-0.0524	0.0301	-0.1517	0.1185
500	0.10	0.02	BS-MID	500	-0.0070	0.0704	-0.0105	-0.0506	0.0368	-0.1510	0.1304
500	0.10	0.05	MI-BS 1	500	-0.0266	0.0882	-0.0249	-0.0861	0.0330	-0.1974	0.1307
500	0.10	0.05	MI-BS 2	500	-0.0276	0.0875	-0.0263	-0.0826	0.0242	-0.1998	0.1417
500	0.10	0.05	BS-MI	500	-0.0200	0.1000	-0.0186	-0.0845	0.0432	-0.2271	0.1715
500	0.10	0.05	MID-BS 1	500	-0.0284	0.0880	-0.0290	-0.0848	0.0287	-0.2007	0.1446
500	0.10	0.05	MID-BS 2	500	-0.0286	0.0854	-0.0266	-0.0827	0.0251	-0.2087	0.1373
500	0.10	0.05	BS-MID	500	-0.0185	0.0896	-0.0135	-0.0756	0.0357	-0.2040	0.1576
500	0.10	0.10	MI-BS 1	500	-0.0338	0.1147	-0.0368	-0.1079	0.0527	-0.2690	0.1633
500	0.10	0.10	MI-BS 2	500	-0.0341	0.1128	-0.0341	-0.1013	0.0467	-0.2627	0.1673
500	0.10	0.10	BS-MI	500	-0.0196	0.1313	-0.0197	-0.1011	0.0667	-0.2662	0.2168
500	0.10	0.10	MID-BS 1	500	-0.0393	0.1121	-0.0408	-0.1131	0.0425	-0.2652	0.1673
500	0.10	0.10	MID-BS 2	500	-0.0376	0.1117	-0.0386	-0.1107	0.0417	-0.2617	0.1706
500	0.10	0.10	BS-MID	500	-0.0162	0.1159	-0.0190	-0.0962	0.0639	-0.2359	0.2057
500	0.25	0.02	MI-BS 1	500	0.0045	0.0529	0.0045	-0.0293	0.0393	-0.1014	0.1025
500	0.25	0.02	MI-BS 2	500	0.0046	0.0522	0.0044	-0.0286	0.0378	-0.1004	0.1077
500	0.25	0.02	BS-MI	500	0.0052	0.0561	0.0062	-0.0343	0.0423	-0.0970	0.1120
500	0.25	0.02	MID-BS 1	500	0.0046	0.0534	0.0040	-0.0276	0.0378	-0.1020	0.1120

Estadísticos descriptivos diferencia  $\Delta$  CITL (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
500	0.25	0.02	MID-BS 2	500	0.0043	0.0517	0.0036	-0.0258	0.0377	-0.1010	0.1053
500	0.25	0.02	BS-MID	500	0.0053	0.0534	0.0055	-0.0301	0.0401	-0.0994	0.1161
500	0.25	0.05	MI-BS 1	500	-0.0012	0.0615	0.0016	-0.0402	0.0425	-0.1267	0.1172
500	0.25	0.05	MI-BS 2	500	-0.0021	0.0595	0.0011	-0.0375	0.0397	-0.1251	0.1095
500	0.25	0.05	BS-MI	500	0.0019	0.0679	0.0039	-0.0402	0.0461	-0.1334	0.1266
500	0.25	0.05	MID-BS 1	500	-0.0033	0.0587	-0.0023	-0.0412	0.0374	-0.1249	0.1037
500	0.25	0.05	MID-BS 2	500	-0.0028	0.0579	-0.0006	-0.0429	0.0349	-0.1224	0.1043
500	0.25	0.05	BS-MID	500	0.0019	0.0605	0.0030	-0.0383	0.0433	-0.1164	0.1092
500	0.25	0.10	MI-BS 1	500	-0.0088	0.0783	-0.0124	-0.0596	0.0480	-0.1621	0.1406
500	0.25	0.10	MI-BS 2	500	-0.0095	0.0775	-0.0117	-0.0606	0.0472	-0.1679	0.1328
500	0.25	0.10	BS-MI	500	-0.0004	0.0927	-0.0058	-0.0590	0.0631	-0.1743	0.1772
500	0.25	0.10	MID-BS 1	500	-0.0114	0.0743	-0.0051	-0.0564	0.0381	-0.1666	0.1247
500	0.25	0.10	MID-BS 2	500	-0.0117	0.0729	-0.0091	-0.0575	0.0372	-0.1548	0.1197
500	0.25	0.10	BS-MID	500	-0.0020	0.0774	0.0020	-0.0548	0.0515	-0.1602	0.1405
500	0.50	0.02	MI-BS 1	500	-0.0032	0.0502	-0.0034	-0.0358	0.0287	-0.0971	0.1013
500	0.50	0.02	MI-BS 2	500	-0.0038	0.0494	-0.0037	-0.0342	0.0243	-0.0999	0.0987
500	0.50	0.02	BS-MI	500	-0.0033	0.0512	-0.0018	-0.0370	0.0285	-0.1044	0.0979
500	0.50	0.02	MID-BS 1	500	-0.0031	0.0493	-0.0019	-0.0337	0.0248	-0.1034	0.0956
500	0.50	0.02	MID-BS 2	500	-0.0031	0.0494	-0.0024	-0.0344	0.0237	-0.1012	0.0996
500	0.50	0.02	BS-MID	500	-0.0022	0.0507	-0.0019	-0.0326	0.0284	-0.1047	0.1028
500	0.50	0.05	MI-BS 1	500	-0.0062	0.0577	-0.0031	-0.0445	0.0303	-0.1115	0.1020
500	0.50	0.05	MI-BS 2	500	-0.0057	0.0569	-0.0024	-0.0439	0.0300	-0.1111	0.1040
500	0.50	0.05	BS-MI	500	-0.0053	0.0605	-0.0049	-0.0410	0.0356	-0.1263	0.0964
500	0.50	0.05	MID-BS 1	500	-0.0057	0.0566	-0.0022	-0.0444	0.0300	-0.1046	0.1037
500	0.50	0.05	MID-BS 2	500	-0.0044	0.0559	-0.0001	-0.0427	0.0309	-0.1059	0.1076
500	0.50	0.05	BS-MID	500	-0.0033	0.0564	0.0016	-0.0429	0.0327	-0.1098	0.1069
500	0.50	0.10	MI-BS 1	500	-0.0108	0.0669	-0.0141	-0.0515	0.0348	-0.1419	0.1171
500	0.50	0.10	MI-BS 2	500	-0.0103	0.0665	-0.0132	-0.0526	0.0331	-0.1394	0.1170
500	0.50	0.10	BS-MI	500	-0.0060	0.0735	-0.0063	-0.0551	0.0408	-0.1451	0.1374
500	0.50	0.10	MID-BS 1	500	-0.0091	0.0665	-0.0116	-0.0545	0.0284	-0.1280	0.1211
500	0.50	0.10	MID-BS 2	500	-0.0100	0.0649	-0.0110	-0.0525	0.0288	-0.1326	0.1216
500	0.50	0.10	BS-MID	500	-0.0043	0.0674	-0.0065	-0.0476	0.0347	-0.1307	0.1338
1000	0.10	0.02	MI-BS 1	500	0.0004	0.0462	-0.0018	-0.0314	0.0318	-0.0872	0.0943
1000	0.10	0.02	MI-BS 2	500	0.0008	0.0460	-0.0033	-0.0287	0.0334	-0.0829	0.0886
1000	0.10	0.02	BS-MI	500	0.0023	0.0514	0.0003	-0.0319	0.0354	-0.0961	0.1042
1000	0.10	0.02	MID-BS 1	500	0.0016	0.0461	0.0005	-0.0299	0.0310	-0.0836	0.0925
1000	0.10	0.02	MID-BS 2	500	0.0009	0.0448	-0.0024	-0.0283	0.0295	-0.0830	0.0868
1000	0.10	0.02	BS-MID	500	0.0033	0.0472	0.0025	-0.0264	0.0327	-0.0863	0.0945
1000	0.10	0.05	MI-BS 1	500	-0.0072	0.0584	-0.0039	-0.0460	0.0317	-0.1332	0.1064

Estadísticos descriptivos diferencia  $\Delta$  CITL (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.10	0.05	MI-BS 2	500	-0.0083	0.0579	-0.0065	-0.0464	0.0304	-0.1287	0.1044
1000	0.10	0.05	BS-MI	500	-0.0033	0.0688	-0.0007	-0.0481	0.0404	-0.1413	0.1318
1000	0.10	0.05	MID-BS 1	500	-0.0094	0.0551	-0.0074	-0.0471	0.0284	-0.1181	0.0999
1000	0.10	0.05	MID-BS 2	500	-0.0093	0.0548	-0.0071	-0.0485	0.0277	-0.1158	0.1015
1000	0.10	0.05	BS-MID	500	-0.0039	0.0578	-0.0023	-0.0436	0.0349	-0.1171	0.1170
1000	0.10	0.10	MI-BS 1	500	-0.0123	0.0766	-0.0169	-0.0654	0.0426	-0.1533	0.1376
1000	0.10	0.10	MI-BS 2	500	-0.0123	0.0762	-0.0141	-0.0655	0.0429	-0.1527	0.1369
1000	0.10	0.10	BS-MI	500	-0.0034	0.0899	-0.0055	-0.0652	0.0526	-0.1791	0.1775
1000	0.10	0.10	MID-BS 1	500	-0.0151	0.0735	-0.0171	-0.0643	0.0325	-0.1634	0.1233
1000	0.10	0.10	MID-BS 2	500	-0.0147	0.0721	-0.0189	-0.0632	0.0337	-0.1504	0.1188
1000	0.10	0.10	BS-MID	500	-0.0026	0.0752	-0.0015	-0.0537	0.0427	-0.1560	0.1420
1000	0.25	0.02	MI-BS 1	500	0.0012	0.0391	-0.0020	-0.0274	0.0299	-0.0659	0.0885
1000	0.25	0.02	MI-BS 2	500	0.0014	0.0380	-0.0019	-0.0254	0.0284	-0.0635	0.0819
1000	0.25	0.02	BS-MI	500	0.0026	0.0402	0.0001	-0.0271	0.0315	-0.0691	0.0850
1000	0.25	0.02	MID-BS 1	500	0.0014	0.0382	-0.0013	-0.0252	0.0270	-0.0681	0.0813
1000	0.25	0.02	MID-BS 2	500	0.0015	0.0375	-0.0019	-0.0244	0.0271	-0.0626	0.0810
1000	0.25	0.02	BS-MID	500	0.0022	0.0383	-0.0009	-0.0251	0.0278	-0.0673	0.0861
1000	0.25	0.05	MI-BS 1	500	-0.0001	0.0440	-0.0028	-0.0280	0.0321	-0.0824	0.0904
1000	0.25	0.05	MI-BS 2	500	-0.0003	0.0427	-0.0036	-0.0279	0.0276	-0.0805	0.0858
1000	0.25	0.05	BS-MI	500	0.0015	0.0483	0.0004	-0.0326	0.0341	-0.0871	0.0989
1000	0.25	0.05	MID-BS 1	500	-0.0004	0.0419	-0.0011	-0.0276	0.0261	-0.0785	0.0824
1000	0.25	0.05	MID-BS 2	500	-0.0012	0.0414	-0.0024	-0.0280	0.0262	-0.0808	0.0833
1000	0.25	0.05	BS-MID	500	0.0013	0.0429	-0.0001	-0.0291	0.0316	-0.0798	0.0845
1000	0.25	0.10	MI-BS 1	500	-0.0031	0.0503	-0.0032	-0.0367	0.0294	-0.1013	0.0980
1000	0.25	0.10	MI-BS 2	500	-0.0028	0.0504	-0.0040	-0.0361	0.0300	-0.1013	0.0966
1000	0.25	0.10	BS-MI	500	0.0014	0.0586	0.0038	-0.0388	0.0394	-0.1147	0.1129
1000	0.25	0.10	MID-BS 1	500	-0.0033	0.0483	-0.0022	-0.0344	0.0280	-0.0956	0.0924
1000	0.25	0.10	MID-BS 2	500	-0.0031	0.0475	-0.0041	-0.0372	0.0255	-0.0909	0.0889
1000	0.25	0.10	BS-MID	500	0.0022	0.0497	0.0030	-0.0301	0.0365	-0.0915	0.0958
1000	0.50	0.02	MI-BS 1	500	-0.0004	0.0364	0.0005	-0.0250	0.0228	-0.0729	0.0672
1000	0.50	0.02	MI-BS 2	500	-0.0001	0.0357	-0.0002	-0.0238	0.0242	-0.0686	0.0674
1000	0.50	0.02	BS-MI	500	-0.0006	0.0370	0.0000	-0.0258	0.0232	-0.0725	0.0709
1000	0.50	0.02	MID-BS 1	500	0.0003	0.0364	0.0014	-0.0234	0.0233	-0.0636	0.0735
1000	0.50	0.02	MID-BS 2	500	0.0002	0.0355	0.0008	-0.0241	0.0246	-0.0665	0.0737
1000	0.50	0.02	BS-MID	500	0.0007	0.0360	0.0025	-0.0253	0.0226	-0.0683	0.0746
1000	0.50	0.05	MI-BS 1	500	-0.0020	0.0388	0.0002	-0.0276	0.0236	-0.0762	0.0763
1000	0.50	0.05	MI-BS 2	500	-0.0016	0.0383	0.0012	-0.0272	0.0237	-0.0776	0.0760
1000	0.50	0.05	BS-MI	500	0.0000	0.0415	0.0030	-0.0278	0.0266	-0.0847	0.0868
1000	0.50	0.05	MID-BS 1	500	-0.0014	0.0382	0.0016	-0.0261	0.0246	-0.0737	0.0775

Estadísticos descriptivos diferencia  $\Delta$  CITL (MAR)

n	propY	propmiss	Estrategia	NSIM	Media	SD	Mediana	Q1	Q3	P2.5	P97.5
1000	0.50	0.05	MID-BS 2	500	-0.0016	0.0380	0.0013	-0.0266	0.0237	-0.0763	0.0731
1000	0.50	0.05	BS-MID	500	-0.0005	0.0392	0.0037	-0.0267	0.0244	-0.0811	0.0772
1000	0.50	0.10	MI-BS 1	500	-0.0048	0.0440	-0.0066	-0.0344	0.0227	-0.0905	0.0842
1000	0.50	0.10	MI-BS 2	500	-0.0050	0.0434	-0.0063	-0.0357	0.0223	-0.0879	0.0815
1000	0.50	0.10	BS-MI	500	-0.0027	0.0485	-0.0031	-0.0330	0.0292	-0.0985	0.0925
1000	0.50	0.10	MID-BS 1	500	-0.0040	0.0430	-0.0053	-0.0321	0.0237	-0.0863	0.0861
1000	0.50	0.10	MID-BS 2	500	-0.0039	0.0429	-0.0062	-0.0325	0.0263	-0.0869	0.0874
1000	0.50	0.10	BS-MID	500	-0.0010	0.0442	-0.0026	-0.0314	0.0260	-0.0866	0.0953

## Anexo 19. Código R para la simulación

```

# Librerías ----
library(parallel)
library(MASS)
library(mice)
library(pROC)
library(dplyr)
library(stringr)

# 1. Generación inicial de datos -----

# n: número de observaciones a generar
# cov.pred: covarianza entre predictores X

gen.data <- function(n, propY, betas, cov.pred = 0){

# Simulación predictores X
n.pred <- length(betas)
mu <- rep(0, n.pred)
var.cov <- matrix(0, nrow = n.pred, ncol = n.pred)
var.cov[1:n.pred, 1:n.pred] <- cov.pred
diag(var.cov) <- 1

x <- mvrnorm(n, mu, var.cov)
colnames(x) <- paste("x", 1:n.pred, sep = "")

# Simulación variable de interés Y
solbeta0 <- function(beta0){
  mean(1 / (1 + exp(-beta0 - x %*% betas))) - propY
}
beta0 <- tryCatch(uniroot(solbeta0, c(-100,100), tol = 1E-7)$root, error = function(e)
NULL)
if(is.null(beta0)){
  beta0 <- mean(log(propY/(1-propY)) - x%*% betas)
}
y <- as.integer(runif(n) < 1 / (1 + exp(-beta0 - x %*% betas)))
list(y = y, x = x)
}

# 2. Selección de muestras completas -----

# NSIM: número de muestras
# n.muestra: tamaño muestral

nsim.muestras <- function(datos.pob, NSIM, n.muestra){

npob <- length(datos.pob[[1]])
propY <- sum(datos.pob[[1]] == 1)/npob

nY <- round(propY * n.muestra)
nnoY <- n.muestra - nY

ind.Y <- which(datos.pob[[1]] == 1)
ind.noY <- setdiff(1:npob, ind.Y)

lapply(1:NSIM, function(nsim){
  ind.Y.muestra <- sample(ind.Y, nY)
  ind.noY.muestra <- sample(ind.noY, nnoY)

```

```

muestra.Y <- datos.pob[[1]][c(ind.Y.muestra, ind.noY.muestra)]
muestra.X <- datos.pob[[2]][c(ind.Y.muestra, ind.noY.muestra), ]

muestra <- list(y = muestra.Y, x = muestra.X)

evaluacion.Y <- datos.pob[[1]][-c(ind.Y.muestra, ind.noY.muestra)]
evaluacion.X <- datos.pob[[2]][-c(ind.Y.muestra, ind.noY.muestra), ]

evaluacion <- list(y = evaluacion.Y, x = evaluacion.X)

list(muestra = muestra, pob.evaluacion = evaluacion)
})
}

# 3. Introducción de valores perdidos (MCAR) ----

MCAR <- function(datos, prop, missY = TRUE){

  MCAR.vect <- function(vector){
    ind <- runif(n = length(vector))
    vector[ind < prop] <- NA
    vector
  }
  if (missY){
    datos.miss <- list(y = MCAR.vect(datos$y), x = apply(datos$x, MARGIN = 2, FUN =
MCAR.vect))
  }
  else {
    datos.miss <- list(y = datos$y, x = apply(datos$x, MARGIN = 2, FUN = MCAR.vect))
  }
  datos.miss
}

# 3. Introducción de valores perdidos (MAR) ----
MAR <- function(datos, prop, missY = TRUE){

  n <- nrow(datos[[2]])
  num.cov <- ncol(datos[[2]])

  Iobs.var <- function(var){
    var[is.na(var)] <- 0
    var
  }

  # Perdidos en Y
  MAR.y <- function(vect){
    betas <- 2

    ind.x.aleat <- sample(1:num.cov, 1)
    x.aleat <- datos[[2]][, ind.x.aleat]
    x <- Iobs.var(x.aleat)

    solbeta0 <- function(beta0){
      mean(1 / (1 + exp(-beta0 - x * betas))) - prop
    }
    beta0 <- tryCatch(uniroot(solbeta0, c(-100,100), tol = 1E-7)$root, error = function(e)
NULL)
    if(is.null(beta0)){
      beta0 <- mean(log(prop/(1-prop))) - x * betas)
    }
  }
}

```

```

    }
    vect[runif(n) < 1 / (1 + exp(-beta0 - x * betas))] <- NA
    vect
  }

# Perdidos en predictores X:
MAR.x <- function(vect, ind.x){

  betas <- c(1, 2, 2)

  ind.x.aleat <- sample(setdiff(1:num.cov, ind.x), 2)
  x.aleat1 <- datos[[2]][, ind.x.aleat[1]]
  x.aleat2 <- datos[[2]][, ind.x.aleat[2]]
  m.aleat1 <- as.numeric(is.na(x.aleat1))

  x.obs <- Iobs.var(x.aleat2)
  y.obs <- Iobs.var(datos[[1]])
  x <- cbind(m.aleat1, x.obs, y.obs)

  solbeta0 <- function(beta0){
    mean(1 / (1 + exp(-beta0 - x %*% betas))) - prop
  }
  beta0 <- tryCatch(uniroot(solbeta0, c(-100,100), tol = 1E-7)$root, error = function(e)
NULL)
  if(is.null(beta0)){
    beta0 <- mean(log(prop/(1-prop)) - x %*% betas)
  }

  vect[runif(n) < 1 / (1 + exp(-beta0 - x %*% betas))] <- NA
  vect
}
list(y = MAR.y(datos[[1]]), x = sapply(1:ncol(datos[[2]]), function(ind)
MAR.x(datos[[2]][,ind], ind.x = ind)))
}

```

#### # 4. Imputación múltiple (MICE) ----

```

MICE <- function(datos.miss, M, usarY = TRUE, pooled = FALSE){

  datos.miss <- cbind(datos.miss$y, datos.miss$x)
  colnames(datos.miss) <- c("y", paste("x", 1:5, sep = ""))

  method = c("Logreg", rep("pmm", ncol(datos.miss)-1))

  predmatrix <- mice(datos.miss, maxit = 0, pri = FALSE)$pred
  if (!usarY){
    predmatrix[, "y"] <- 0
  }
  IMP <- mice(data = datos.miss, m = M, method = method, predictorMatrix = predmatrix,
printFlag = F)

# MI:
listaMI <- lapply(1:M, function(m) {
  data.frame <- complete(IMP, m)
  lista <- list(y = data.frame[,1], x = as.matrix(data.frame[, -1]))
  lista})

# MID:
ind.Yobs <- !is.na(datos.miss[,1])

```

```

ListaMID <- Lapply(listaMI, FUN = function(muestra.imp) {
  list(y = muestra.imp[[1]][ind.Yobs], x = as.matrix(muestra.imp[[2]][ind.Yobs, ]))})

if(pooled){
# Modelo pooled (MI)
modelos.mi <- with(data = IMP, expr = glm(y ~ x1 + x2 + x3 + x4 + x5, family = binomial))
pooled.mi <- summary(pool(modelos.mi))$estimate

# Modelo pooled (MID)
modelos.mid <- sapply(ListaMID, function(datos.mid){
  modelo <- glm(y ~ x, family = binomial, data = datos.mid)
  modelo$coefficients
})
pooled.mid <- unname(rowMeans(modelos.mid))

# Modelos pooled glm (MI y MID)
pooled.mi.glm <- modelos.mi$analyses[[1]]
pooled.mid.glm <- modelos.mi$analyses[[1]]

pooled.mi.glm$coefficients <- pooled.mi
pooled.mid.glm$coefficients <- pooled.mid

list(MI = listaMI, MID = listaMID, pooled.mi = pooled.mi.glm, pooled.mid =
pooled.mid.glm)
}
else
list(MI = listaMI, MID = listaMID)
}

# 4. Remuestreo bootstrap ----
BS <- function(datos, B, ind = FALSE){
  n <- length(datos$y)
  BS.uno <- function(b){
    ind.bs <- sample(1:n, n, replace = TRUE)
    if (ind){
      result <- ind.bs
    }
    else {
      result <- list(y = datos[[1]][ind.bs], x = datos[[2]][ind.bs, ])
    }
  }
  result
}
lista <- Lapply(1:B, BS.uno)
}

# 5. Rendimiento predictivo ----
rendimiento <- function(datos.ajuste, datos.eval, validacion = TRUE){
  medidas <- c("auc", "cal.citl", "cal.slope")
  nombres.ap <- paste(medidas, ".ap", sep = "")
  nombres.test <- paste(medidas, ".test", sep = "")

# Rendimiento aparente
y.ajuste <- datos.ajuste[[1]]
modelo <- glm(y ~ x, family = binomial, data = datos.ajuste)
prob.pred.ajuste <- predict(modelo, type = "response")
lp.pred.ajuste <- predict(modelo)

# Estadístico C
auc <- pROC::roc(response = y.ajuste, predictor = prob.pred.ajuste, direction = "<")$auc

# CITL y calibration slope
citl <- 0

```

```

slope <- 1

rend <- c(auc, citl, slope)
names(rend) <- nombres.ap

# Rendimiento test
if (validacion){
  y.eval <- datos.eval[[1]]

  prob.pred.eval <- stats::predict(modelo, type = "response", newdata = datos.eval)
  lp.pred.eval <- stats::predict(modelo, newdata = datos.eval)

  # Estadístico C
  auc.test <- pROC::roc(response = y.eval, predictor = prob.pred.eval, direction =
"<")$auc
  # CITL y calibration slope
  citl.eval <- summary(glm(y.eval ~ offset(lp.pred.eval), family = "binomial"))$coef[1,1]
  slope.eval <- summary(glm(y.eval ~ lp.pred.eval, family = "binomial"))$coef[2,1]

  rend <- c(rend, auc.test, citl.eval, slope.eval)
  names(rend) <- c(nombres.ap, nombres.test)
}
rend
}

rendimiento.error <- function(datos.ajuste, datos.eval, validacion = TRUE){
  medidas <- c("auc", "cal.citl", "cal.slope")
  nombres.ap <- paste(medidas, ".ap", sep = "")
  nombres.test <- paste(medidas, ".test", sep = "")
  tryCatch(rendimiento(datos.ajuste, datos.eval, validacion), error = function(e){
    rend <- rep(NA, length(medidas))
    names(rend) <- nombres.ap
    if(validacion){
      rend <- c(rend, rend)
      names(rend) <- c(nombres.ap, nombres.test)
    }
    rend
  })
}

rendimiento.pooled <- function(modelo.pooled, datos.eval){
  medidas <- c("auc", "cal.citl", "cal.slope")
  nombres.pob <- paste(medidas, "pob", sep = ".")

  y.eval <- datos.eval[[1]]
  datos.eval <- as.data.frame(cbind(datos.eval$y, datos.eval$x))
  names(datos.eval)[1] <- "y"
  prob.pred.eval <- stats::predict(modelo.pooled, type = "response", newdata = datos.eval)
  lp.pred.eval <- stats::predict(modelo.pooled, newdata = datos.eval)

  # Estadístico C
  auc.eval <- pROC::roc(response = y.eval, predictor = prob.pred.eval, direction = "<")$auc
  # CITL y calibration slope
  citl.eval <- summary(glm(y.eval ~ offset(lp.pred.eval), family = "binomial"))$coef[1,1]
  slope.eval <- summary(glm(y.eval ~ lp.pred.eval, family = "binomial"))$coef[2,1]

  rend <- c(auc.eval, citl.eval, slope.eval)
  names(rend) <- nombres.pob
  rend
}

rendimiento.pooled.error <- function(modelo.pooled, datos.eval){

```

```

medidas <- c("auc", "cal.citl", "cal.slope")
nombres.pob <- paste(medidas, "pob", sep = ".")

tryCatch(rendimiento.pooled(modelo.pooled, datos.eval), error = function(e){
  rend <- rep(NA, length(medidas))
  names(rend) <- nombres.pob
  rend
})
}

# 6. Optimismo ----

# estrategia: "MICE-BS", "BS-MICE"

optimismo.rend <- function(MBmuestras, datos.eval, estrategia){
  medidas <- c("auc", "cal.citl", "cal.slope")

  if (estrategia == "MICE-BS"){
    rendMB <- lapply(1:length(MBmuestras), function(m)
      t(sapply(1:length(MBmuestras[[1]]), function(b) {
        rendimiento.error(datos.ajuste = MBmuestras[[m]][[b]], datos.eval =
datos.eval[[m]]})))
  }
  else{
    rendMB <- lapply(1:length(MBmuestras), function(b)
      t(sapply(1:length(MBmuestras[[1]]), function(m) {
        rendimiento.error(datos.ajuste = MBmuestras[[b]][[m]], datos.eval =
datos.eval[[m]]})))
  }
  rendMB <- do.call(rbind, rendMB)
  ind.ap <- 1:length(medidas)
  rendMB.dif <- rendMB[, ind.ap] - rendMB[, -ind.ap]
  rendMB.opt <- colMeans(rendMB.dif, na.rm = TRUE)
  rendMB.opt
}

# 7. Simulación ----
simulacion.uno <- function(datos.muestra, datos.test, n, propY, betas, cov.pred,
  prop.miss, patron.miss = "MCAR"){

  medidas <- c("auc", "cal.citl", "cal.slope")

  # Introducción valores perdidos
  if (patron.miss == "MCAR"){
    datos.miss <- MCAR(datos.muestra, prop.miss)
  }
  else{
    datos.miss <- MAR(datos.muestra, prop.miss)
  }

  # Casos completos (se descartan observaciones con valores perdidos)
  ind.comp <- complete.cases(datos.miss)
  datos.casoscomp <- list(y = datos.muestra[[1]][ind.comp], x =
datos.muestra[[2]][ind.comp, ])

  # Imputación multiple MICE
  datos.imp <- MICE(datos.miss = datos.miss, M = M, pooled = TRUE)
  # MI ----
  datos.mi <- datos.imp[[1]]

```

```

# MID ----
datos.mid <- datos.imp[[2]]

# Muestras BS sin perdidos
BS.comp <- BS(datos.muestra, B)
# Muestras BS casos completos
BS.casoscomp <- BS(datos.casoscomp, B)

# Rendimiento sin perdidos
ind.ap <- 1:length(medidas)
rend.sinperd <- rendimiento.error(datos.muestra, datos.test)

# De referencia
rend.pob.sinperd <- rend.sinperd[-ind.ap]
# Aparente
r.ap.comp <- rend.sinperd[ind.ap]
# BS
r.bs.comp <- t(sapply(BS.comp, rendimiento.error, datos.eval = datos.muestra))
r.bs.dif <- r.bs.comp[, ind.ap] - r.bs.comp[, -ind.ap]
r.bs.comp.opt <- colMeans(r.bs.dif, na.rm = TRUE)

# Rendimiento casos completos
rend.casoscomp <- rendimiento.error(datos.casoscomp, datos.test)
# De referencia
rend.pob.casoscomp <- rend.casoscomp[-ind.ap]
# Aparente
r.ap.casoscomp <- rend.casoscomp[ind.ap]
# BS
r.bs.casoscomp <- t(sapply(BS.casoscomp, rendimiento.error, datos.eval =
datos.casoscomp))
r.bs.casoscomp.dif <- r.bs.casoscomp[, ind.ap] - r.bs.casoscomp[, -ind.ap]
r.bs.casoscomp.opt <- colMeans(r.bs.casoscomp.dif, na.rm = TRUE)

# BS con valores perdidos
t.ini <- Sys.time()
BS.miss <- BS(datos.miss, B)
t.fin <- Sys.time()
t.bsmiss <- t.fin - t.ini
rm(t.ini, t.fin)

t.ini <- Sys.time()
BSMICE <- lapply(BS.miss, MICE, M = M)
t.fin <- Sys.time()
t.bsmiss <- t.fin - t.ini
rm(t.ini, t.fin)

# MI

# Rendimiento MI
# De refernecia (modelo pooled)
modelo.pooled.mi <- datos.imp$pooled.mi
rend.pob.mi <- rendimiento.pooled.error(modelo.pooled.mi, datos.test)
# Aparente
r.ap.mi <- t(sapply(1:length(datos.mi), function(m) rendimiento.error(datos.mi[[m]],
datos.mi[[m]], validacion = FALSE)))
r.ap.mi.medio <- colMeans(r.ap.mi, na.rm = TRUE)

# MI-BS 1
t.ini <- Sys.time()
indMI.BS <- BS(datos.mi[[1]], B, ind = TRUE)
MIBS1 <- lapply(1:length(datos.mi), function(m) {lapply(1:length(indMI.BS), function(b) {

```

```

    list(y = datos.mi[[m]][[1]][indMI.BS[[b]]], x = datos.mi[[m]][[2]][indMI.BS[[b]], ]))})

r.MIBS1.opt <- optimismo.rend(MIBS1, datos.mi, "MICE-BS")
t.fin <- Sys.time()
t.mibs1 <- t.fin - t.ini
rm(t.fin, t.ini)

# MI-BS 2
t.ini <- Sys.time()
MIBS2 <- lapply(datos.mi, BS, B = B)
r.MIBS2.opt <- optimismo.rend(MIBS2, datos.mi, "MICE-BS")
t.fin <- Sys.time()
t.mibs2 <- t.fin - t.ini
rm(t.fin, t.ini)

# BS-MI
BSMI <- lapply(BSMICE, `[`, 1)
t.ini <- Sys.time()
r.BSMI.opt <- optimismo.rend(BSMI, datos.mi, "BS-MICE")
t.fin <- Sys.time()
t.bsmi <- t.bsmi + t.bsmice + (t.fin - t.ini)
rm(t.fin, t.ini)

# MID
  # De referencia (modelo pooled)
modelo.pooled.mid <- datos.imp$pooled.mid
rend.pob.mid <- rendimiento.pooled.error(modelo.pooled.mid, datos.test)

  # Aparente
r.ap.mid <- t(sapply(1:length(datos.mid), function(m) rendimiento.error(datos.mid[[m]],
datos.mid[[m]], validacion = FALSE)))
r.ap.mid.medio <- colMeans(r.ap.mid, na.rm = TRUE)

  # MID-BS 1
t.ini <- Sys.time()
indMID.BS <- BS(datos.mid[[1]], B, ind = TRUE) MIDBS1 <- lapply(1:length(datos.mid),
function(m) {lapply(1:length(indMID.BS), function(b) {
  list(y = datos.mid[[m]][[1]][indMID.BS[[b]]], x =
datos.mid[[m]][[2]][indMID.BS[[b]], ]))})})
r.MIDBS1.opt <- optimismo.rend(MIDBS1, datos.mid, "MICE-BS")
t.fin <- Sys.time()
t.midbs1 <- t.fin - t.ini
rm(t.fin, t.ini)

# MID-BS 2
t.ini <- Sys.time()
MIDBS2 <- lapply(datos.mid, BS, B = B)
r.MIDBS2.opt <- optimismo.rend(MIDBS2, datos.mid, "MICE-BS")
t.fin <- Sys.time()
t.midbs2 <- t.fin - t.ini
rm(t.fin, t.ini)

# BS-MID
BSMID <- lapply(BSMICE, `[`, 2) # Lista de B listas de M dataframes
t.ini <- Sys.time()
r.BSMID.opt <- optimismo.rend(BSMID, datos.mid, "BS-MICE")
t.fin <- Sys.time()
t.bsmid <- t.bsmi + t.bsmice + (t.fin - t.ini)

# Almacenamiento resultados
estrategias <- c("aparente.comp", "BS.comp",

```

```

    "aparente.casoscomp", "BS.casoscomp",
    "aparente.MI",
    "MI-BS(1)", "MI-BS(2)", "BS-MI",
    "aparente.MID",
    "MID-BS(1)", "MID-BS(2)", "BS-MID")

medidas.ap <- paste(medidas, "ap", sep = ".")
medidas.corr <- paste(medidas, "corr", sep = ".")
medidas.opt <- paste(medidas, "opt", sep = ".")
medidas.ref <- paste(medidas, "ref", sep = ".")
medidas.difref <- paste(medidas, "difref", sep = ".")

result <- data.frame(matrix(nrow = Length(estrategias), ncol = 1 + Length(medidas)*5 +
4))
names(result) <- c("estrategia", medidas.ap, medidas.corr, medidas.opt, medidas.ref,
medidas.difref,
                "n", "propY", "prop.miss", "mecanismo.miss")

result[, "estrategia"] <- estrategias

# Rendimiento aparente
result[result$estrategia %in% c("aparente.comp", "BS.comp"), medidas.ap] <-
rep(r.ap.comp, each = 2)
result[result$estrategia %in% c("aparente.casoscomp", "BS.casoscomp"), medidas.ap] <-
rep(r.ap.casoscomp, each = 2)
result[result$estrategia %in% c("aparente.MI", "MI-BS(1)", "MI-BS(2)", "BS-MI"),
medidas.ap] <- rep(r.ap.mi.medio, each = 4)
result[result$estrategia %in% c("aparente.MID", "MID-BS(1)", "MID-BS(2)", "BS-MID"),
medidas.ap] <- rep(r.ap.mid.medio, each = 4)

# Optimismo en rendimiento aparente
result[result$estrategia %in% c("aparente.comp", "aparente.casoscomp",
"aparente.MI", "aparente.MID"), medidas.corr] <- NA
result[result$estrategia == ("BS.comp"), medidas.opt] <- r.bs.comp.opt
result[result$estrategia == ("BS.casoscomp"), medidas.opt] <- r.bs.casoscomp.opt

result[result$estrategia == ("MI-BS(1)"), medidas.opt] <- r.MIBS1.opt
result[result$estrategia == ("MI-BS(2)"), medidas.opt] <- r.MIBS2.opt
result[result$estrategia == ("BS-MI"), medidas.opt] <- r.BSMI.opt

result[result$estrategia == ("MID-BS(1)"), medidas.opt] <- r.MIDBS1.opt
result[result$estrategia == ("MID-BS(2)"), medidas.opt] <- r.MIDBS2.opt
result[result$estrategia == ("BS-MID"), medidas.opt] <- r.BSMID.opt

# Rendimiento corregido
result[, medidas.corr] <- result[, medidas.ap] - result[, medidas.opt]

# Rendimiento de referencia
result[result$estrategia %in% c("aparente.comp", "BS.comp"), medidas.ref] <-
rep(rend.pob.sinperd, each = 2)
result[result$estrategia %in% c("aparente.casoscomp", "BS.casoscomp"), medidas.ref] <-
rep(rend.pob.casoscomp, each = 2)
result[result$estrategia %in% c("aparente.MI", "MI-BS(1)", "MI-BS(2)", "BS-MI"),
medidas.ref] <- rep(rend.pob.mi, each = 4)
result[result$estrategia %in% c("aparente.MID", "MID-BS(1)", "MID-BS(2)", "BS-MID"),
medidas.ref] <- rep(rend.pob.mid, each = 4)

# Diferencia corregido - referencia
es.aparente <- str_detect(result$estrategia, "aparente")

```

```

# Para rendimientos aparentes: aparente - referencia
result[es.aparente, medidas.difref] <- result[es.aparente, medidas.ap] -
result[es.aparente, medidas.ref]
# Para rendimientos corregidos: corregido - referencia
result[!es.aparente, medidas.difref] <- result[!es.aparente, medidas.corr] -
result[!es.aparente, medidas.ref]

# Características muestra simulada
result[, "n"] <- n
result[, "propY"] <- propY
result[, "prop.miss"] <- prop.miss
result[, "mecanismo.miss"] <- patron.miss

# Resultados rendimiento de referencia (guardados a parte)
modelo.ref <- c("muestra completa", "casos completos", "pooled MI", "pooled MID")
rend.ref <- data.frame(matrix(nrow = length(modelo.ref), ncol = 5 + length(medidas)))
names(rend.ref) <- c("modelo.ref", medidas.ref, "n", "propY", "prop.miss",
"mecanismo.miss")
rend.ref[, "modelo.ref"] <- modelo.ref
rend.ref[, medidas.ref] <- rbind(rend.pob.sinperd, rend.pob.casoscomp, rend.pob.mi,
rend.pob.mid)
rend.ref[, "n"] <- n
rend.ref[, "propY"] <- propY
rend.ref[, "prop.miss"] <- prop.miss
rend.ref[, "mecanismo.miss"] <- patron.miss

rownames(rend.ref) <- NULL

# Resultados tiempos de ejecución
tiempos <- data.frame(matrix(nrow = 6, ncol = 6))
names(tiempos) <- c("estrategia", "tiempo", "n", "propY", "prop.miss", "mecanismo.miss")
tiempos[, "estrategia"] <- c("MI-BS(1)", "MI-BS(2)", "BS-MI", "MID-BS(1)", "MID-BS(2)",
"BS-MID")
tiempos[, "tiempo"] <- c(t.mibs1, t.mibs2, t.bsmi, t.midbs1, t.midbs2, t.bsmid)
tiempos[, "n"] <- n
tiempos[, "propY"] <- propY
tiempos[, "prop.miss"] <- prop.miss
tiempos[, "mecanismo.miss"] <- patron.miss

lista <- list(resultados = result, rendimiento.ref = rend.ref, tiempos = tiempos)
lista
}

simulacion.escenario.parall <- function(datos, n, propY, betas, cov.pred, prop.miss,
patron.miss){
  numcores <- detectCores()
  resul <- mclapply(1:length(datos), FUN = function(nsim)
    simulacion.uno(datos.muestra = datos[[nsim]][[1]], datos.test = datos[[nsim]][[2]], n =
n, propY = propY, betas = betas, cov.pred = cov.pred,
prop.miss = prop.miss, patron.miss = patron.miss), mc.cores = numcores)
  resul
}

# Variables globales
set.seed(18500581)
NSIM <- 500
M <- 5
B <- 100
betas <- c(0.50, 0.50, 0.40, 0.30, 0.30)
cov.pred <- 0.1
N <- 100000

```

```

# Escenarios:
propY <- c(0.10, 0.25, 0.50)
n <- c(100, 500, 1000)
prop.miss <- c(0.02, 0.05, 0.10)
mecanismo.miss <- c("MCAR", "MAR")

# Ejecución
datos.pob <- Lapply(1:length(propY),
                    function(Yprop) gen.data(n = N, propY = propY[Yprop], betas = betas,
cov.pred = cov.pred))

resultados <- Lapply(1:length(propY), function(prop)
  Lapply(1:length(n), function(n.muestra)
    datos.sim <- nsim.muestras(datos.pob[[prop]], NSIM, n[n.muestra])
    result <- Lapply(1:length(mecanismo.miss), function(m.miss)
      Lapply(1:length(prop.miss), function(p.miss)
        simulacion.escenario.parall(datos = datos.sim, n = n[n.muestra], propY =
propY[prop], betas = betas, cov.pred = cov.pred, prop.miss = prop.miss[p.miss], patron.miss
= mecanismo.miss[m.miss])))
      List(result = result)
    )))
  )))

```

