



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Análisis comparativo entre el ajuste por regresión lineal múltiple y Propensity Score ordinal basado en ponderación media Marginal por estratificación (MMWS) en el control de la confusión en estudios observacionales

Sergio Mauricio Moreno López

Universidad Nacional de Colombia
Facultad de Medicina
Instituto de Investigaciones Clínicas
Bogotá, Colombia
2017

Análisis comparativo entre el ajuste por regresión lineal múltiple y Propensity Score ordinal basado en ponderación media Marginal por estratificación (MMWS) en el control de la confusión en estudios observacionales

Sergio Mauricio Moreno López

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Magister en Epidemiología Clínica

Director (a):

MD, MSc. Oscar Alexander Guevara Cruz

Línea de Investigación:

Análisis de simulación

Universidad Nacional de Colombia

Facultad de Medicina

Bogotá, Colombia

2017

Agradecimientos

Principalmente a Dios por darme la oportunidad de culminar esta nueva fase académica, por su acompañamiento y guía en esta etapa de mi vida. A mi familia por su comprensión y apoyo incondicional a lo largo de este proceso. A la Dra. Eliana Nájera por su compañía, respaldo y cariño, su constante aliento fue fundamental para terminar este proyecto.

Agradecimiento especial a todos los docentes de la Maestría en Epidemiología Clínica por sus enseñanzas, al Dr. Oscar Guevara por dirigir este proyecto lado, por su orientación, sugerencias y apoyo en la culminación de este trabajo. Gracias a Janeth Rueda por su acompañamiento y guía en las gestiones académicas y administrativas que culmina con la presentación de este manuscrito. Por ultimo un agradecimiento muy especial a la Profesora Luz Mery González y al Profesor Fabián Armando Gil, quien hicieron importantes recomendaciones al actual documento enriqueciendo tanto la parte metodológica como la sección de resultados de este manuscrito.

A mis compañeros de la Maestría en Epidemiología Clínica por sus enseñanzas a nivel académico y personal

Resumen

Introducción: Desde la perspectiva epidemiológica, existen dos metodologías que permiten determinar la asociación entre variables o factores biológicos: estudios experimentales o no experimentales (1). Se tiene a los estudios epidemiológicos experimentales y específicamente al ensayo clínico como "la herramienta más definitiva para la evaluación de la aplicabilidad de la investigación clínica" (2). Cuando no es factible desarrollar un ensayo clínico aleatorizado, es menester tener en cuenta el fenómeno de la confusión; no ajustar el efecto de un determinado tratamiento configura un sesgo en el análisis que puede invalidar los resultados encontrados (2). Entre los métodos de ajuste más comunes está el modelo de regresión convencional (3,4) y el uso del Propensity score (4,5). En este estudio se busca determinar bajo distintos escenarios de tamaño de muestra, número de variables predictoras y razón expuesto - no expuesto, cual metodología produce los mejores resultados en términos de precisión del estimador del efecto de una exposición o intervención establecida, usando un estimador propuesto para tratamientos ordinales conocido como MMWS y una alternativa de este basado en estimadores doblemente robustos (DR-MMWS).

Métodos: Simulación de Monte Carlo se usa para comparar el estimador MMWS con el ajuste de regresión convencional. Se plantearon diversos escenarios en cuanto al de tamaño de muestra, número de variables predictoras y razón expuesto - no expuesto para evaluar el rendimiento de estos modelos.

Resultados: Los estimadores basados en Propensity Score ordinal basado en ponderación media marginal por estratificación (MMWS), son ligeramente inferiores a la regresión lineal múltiple en la estimación insesgada del efecto en una intervención o exposición de tipo ordinal. El estimador DR-MMWS equipara a los estimadores de una regresión por mínimos cuadrados ordinarios para tamaños de muestra superiores a 500 sujetos. Aumentar el número de no expuestos por expuestos, aumenta la precisión de los estimadores independiente de la metodología de ajuste.

Conclusiones: Debe considerarse el uso de estimadores DR-MMWS en estudios observacionales, ya que, en comparación con los estimadores dados por una regresión lineal múltiple, este estimador presenta iguales indicadores de desempeño y las estimaciones dadas por esta metodología son consistentes y presentan una reducción del sesgo importante

Palabras Clave: Estudio epidemiológico, confusión, relación causal, Propensity score, modelo lineal generalizado, MMWS.

Abstract

Introduction: From the epidemiological perspective, there are two methodologies that allow to determine the association between variables or biological factors: experimental or non-experimental studies (1). We have experimental epidemiological studies and specifically the clinical trial as "the most definitive tool for the evaluation of the applicability of clinical research" (2). Where it is not feasible to develop a randomized clinical trial, it is necessary to take into account the phenomenon of the confusion, not adjusting the effect of a given treatment creates a bias in the analysis that can invalidate the results found. The most common adjustment methods are the conventional regression model (3,4) and the use of the Propensity score (4,5). This study seeks to determine under different scenarios of sample size, number of predictor variables and exposed ratio - not exposed, which methodology produces the best results in terms of the estimator's precision of the effect of an exposure or intervention established, using additionally a proposed estimator for ordinal treatments known as MMWS.

Methods: Monte Carlo simulation is used to compare the MMWS estimator with the conventional regression adjustment. Several scenarios were raised regarding the sample size, the number of prediction variables and the reason why it is exposed, not exposed to evaluate the performance of these models.

Results: The estimators based on ordinal Propensity Score based on marginal average weighting by stratification (MMWS), are slightly lower than the multiple linear regression in the unbiased estimate of the effect in an intervention or ordinal exposure. The DR-MMWS estimator equates the estimators of an ordinary least squares regression for sample sizes greater than 500 subjects. Increasing the number of unexposed by exposed, increases the accuracy of the estimators independent of the adjustment methodology.

Conclusions: The use of DR-MMWS estimators in observational studies should be considered, since, in comparison with the estimators given by a multiple linear regression,

X Análisis comparativo entre el ajuste por regresión lineal múltiple y Propensity Score ordinal basado en ponderación media Marginal por estratificación (MMWS)

this estimator presents the same performance indicators and the estimations given by this methodology are consistent and show a reduction of the bias important

Keywords: epidemiological study, confusion, causal relationship, propensity score, generalized linear model, MMWS.

Contenido

	Pág.
Resumen	VII
Lista de figuras	XIII
Lista de tablas	XV
Lista de Símbolos y abreviaturas	XVI
Introducción	1
1. Justificación y planteamiento del problema	3
2. Pregunta de investigación	5
3. Marco teórico	7
3.1 Conceptos estadísticos	7
3.1.1 Modelos Lineales.....	7
3.1.2 Regresión logística	10
3.1.3 Función Kernel	12
3.1.4 Propensity Score Matching	13
3.1.5 Pasos para la estimación de un Propensity Score	15
3.1.6 Método de Ponderación Media Marginal Basado En Un Propensity Score a través De estratificación (MMWS)	19
3.1.7 Estimadores doblemente robustos.....	21
3.1.8 Métodos de comparación de modelos	23
4. Objetivos	25
4.1 Objetivo general	25
4.2 Objetivos específicos	25
5. Metodología	27
5.1 Tipo de investigación y diseño.....	27
5.2 Hipótesis	27
5.3 Análisis comparativo y de sensibilidad	28
5.3.1 Relación expuestos-no expuestos	28
5.3.2 Escenarios de tamaños de muestra.....	28
5.3.3 Número de variables a incluir.....	29
5.3.4 Análisis de la información	29

5.4	Proceso de simulación del estudio	29
5.5	Proceso de generación de datos para el modelo de tratamiento.....	30
5.6	Pre procesamiento de los datos.....	31
5.7	Consideraciones éticas.....	31
5.8	Financiación.....	31
6.	Resultados	33
6.1	Resultados de la simulación	33
6.1.1	Comparación entre regresión lineal y Propensity Score combinado con MMWS.....	33
6.1.2	Propiedades de los estimadores doblemente robustos (DR-MMWS)	41
6.1.3	Efecto de un mayor número de variables	49
7.	Discusión.....	56
8.	Conclusiones y recomendaciones	61
8.1	Conclusiones	61
8.2	Recomendaciones	62
9.	Bibliografía	63
Anexos.....		68
A.	Aprobación Comité de Ética	68
B.	Implementación de la simulación en R	70
C.	Indicadores de desempeño para el tratamiento/exposición 1	79
D.	Indicadores de desempeño para el tratamiento/exposición 2	81
E.	Evaluación de soporte común y balance	82

Lista de figuras

	Pág.
Figura 3-1: Pasos para la implementación para un Propensity Score	15
Figura 3-2: Tipos de emparejamiento	17
Figura 6-1: Estimaciones del valor del nivel de tratamiento/exposición 1 (MMWS vs MLR)	33
Figura 6-2: Error estándar de la estimación para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)	34
Figura 6-3: Cuadrado Medio del Error para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)	34
Figura 6-4: Error Porcentual Absoluto Medio para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)	35
Figura 6-5: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)	35
Figura 6-6: Valor p observado para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)	36
Figura 6-7: Curva de potencia para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)	36
Figura 6-8: Estimaciones del valor del nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	37
Figura 6-9: Error estándar de la estimación para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	38
Figura 6-10: Cuadrado Medio del Error para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	38
Figura 6-11: Error Porcentual Absoluto Medio para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	39
Figura 6-12: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	39
Figura 6-13: Valor p observado para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	40
Figura 6-14: Curva de potencia para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)	40
Figura 6-15: Estimaciones del valor del nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)	41

Figura 6-16: Error estándar de la estimación para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)	42
Figura 6-17: Cuadrado Medio del Error del nivel de tratamiento/exposición 1 (DR-MMWS vs MLR).....	42
Figura 6-18: Error Porcentual Absoluto Medio del nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)	43
Figura 6-19: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)	43
Figura 6-20: Valor p observado para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)	44
Figura 6-21: Curva de potencia para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)	44
Figura 6-22: Estimaciones del valor del nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)	45
Figura 6-23: Error estándar de la estimación para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)	46
Figura 6-24: Cuadrado Medio del Error del nivel de tratamiento/exposición 2 (DR-MMWS vs MLR).....	46
Figura 6-25: Error Porcentual Absoluto Medio del nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)	47
Figura 6-26: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)	47
Figura 6-27: Valor p observado para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)	48
Figura 6-28: Curva de potencia para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)	48
Figura 6-29: Comparación de las estimaciones del valor del nivel de tratamiento/ 1 (DR-MMWS, MLR y MMWS PS).....	49
Figura 6-30: Comparación de los Errores estándar de la estimación para el nivel de tratamiento/exposición 1 (DR-MMWS, MLR y MMWS PS).....	50
Figura 6-31: Comparación de las estimaciones del valor del nivel de tratamiento/ 2 (DR-MMWS, MLR y MMWS PS).....	53
Figura 6-32: Comparación de los Errores estándar de la estimación para el nivel de tratamiento/exposición 1 (DR-MMWS, MLR y MMWS PS).....	53

Lista de tablas

	Pág.
Tabla 3-1: Principales funciones Kernel	12
Tabla 5-1: Escenarios de tamaño de muestra establecidos para el análisis	28
Tabla 6-1: Resultados de la simulación de Monte Carlo para los métodos comparados con diversos números de variables predictoras para el tratamiento/exposición 1	51
Tabla 6-2: Resultados de la simulación de Monte Carlo para los métodos comparados con diversos números de variables predictoras para el tratamiento/exposición 2	54

Lista de Símbolos y abreviaturas

Abreviatura	Término
ACE	Efecto Causal Promedio
H_0	Hipótesis Nula
H_a	Hipótesis Alternativa
IC	Intervalo de Confianza
MEM	Modelos Estructurales Marginales
ML	Modelos lineales
MLG	Modelos Lineales Generalizados
PS	Propensity Score
RIC	Rango Intercuartílico
MMWS	Marginal Mean Weighting through Stratification
MAPE	Error Porcentual Absoluto Medio (Mean Absolute Percentage Error)
MSE	Cuadrado Medio del Error (Mean Square Error)
SMD	Diferencia Promedio Estandarizada (Standardized Mean Difference)

Introducción

Dentro de la investigación clínica y epidemiológica, existen dos metodologías que permiten determinar la asociación entre variables o factores biológicos: estudios experimentales o no experimentales (1). Se tiene a los estudios epidemiológicos experimentales y específicamente al ensayo clínico como "la herramienta más definitiva para la evaluación de la aplicabilidad de la investigación clínica" (1). Representa "una actividad de investigación clave con el potencial de mejorar la calidad de la atención de la salud y los costos de control a través de una cuidadosa comparación de tratamientos alternativos" (2). En los ensayos clínicos, la asignación aleatoria de los pacientes a los grupos de intervención o de control busca el equilibrio de estos grupos en cuanto a esas variables que afectan la asociación entre la exposición y el efecto (1). Esto garantiza que las estimaciones del efecto de un tratamiento o exposición sean insesgadas.

Aunque los ensayos clínicos son el referente en cuanto a investigación clínica, factores éticos y de factibilidad pueden dificultar la implementación de este tipo de estudios. Ante esta eventualidad los estudios no experimentales u observacionales, son una alternativa para determinar estimaciones del efecto (1,2,6), dado que el investigador observa el fenómeno tal y como se presenta, sin poder controlar las variables que lo modifican, atenúan o acentúan, sean estas conocidas o desconocidas por lo que estos estudios son propensos a presentar resultados que "difieren sistemáticamente de la verdad" debido a procedimientos inadecuados en cualquier etapa de un estudio (7). Esto último es muy importante si se busca estimaciones no sesgadas del efecto de una exposición o intervención en un desenlace, ya que se debe estratificar o emparejar los sujetos con el fin de que los sujetos de los grupos de tratamiento y control de cada estrato o pareja estén equilibrados con respecto a las covariables de interés (2,8).

Estos procedimientos mencionados anteriormente son sencillos cuando se tiene una covariable sospechosa de causar confusión entre la asociación de una exposición con un desenlace (9). De no ser así, la estratificación o emparejamiento es dispendioso, dificulta el proceso de recolección de la muestra, disminuye la cantidad de información a analizar y

disminuye la probabilidad de una verdadera asociación entre la exposición y el desenlace (9,10). Este tipo de escenarios conlleva al uso de métodos multivariados, siendo uno de los más usados en el contexto epidemiológico los modelos de regresión(8), no obstante, ante eventos pocos frecuentes o cuando el número de covariables es muy amplio con respecto al tamaño de la muestra, los estimadores son inexactos y tienen problemas de precisión (10,11).

Ante estas dificultades, Rosenbaum y Rubin propusieron una metodología de análisis conocida como Propensity Score (PS), que se define como la probabilidad condicional de ser asignado a un tratamiento o exposición particular dado un conjunto de covariables observadas antes de recibir el tratamiento (12–14). Según la literatura tanto en tamaños de muestra grandes como pequeños, el ajuste por el Propensity Score es suficiente para reducir los sesgos debido al efecto dado por covariables observadas (15), adicionalmente esta propuesta corrige la estimación del efecto de un tratamiento en diseños no experimentales de manera integral basado en la idea de que el sesgo se reduce cuando la comparación de los resultados se realiza entre sujetos tratados y no tratados lo más similares posibles en cuanto a las variables observadas (16–20).

1. Justificación y planteamiento del problema

Aunque existen en la literatura estudios basados en análisis comparativos entre distintos modelos lineales y el Propensity Score que han determinado la ventaja de este último en algunos escenarios, estos solo analizan una intervención de tipo dicótoma, lo cual no permite evaluar el efecto de los niveles de tratamiento en el desenlace de interés. Zannuto en su estudio del 2006 (21) buscaba determinar la posible existencia de diferencias entre los estimadores de una Regresión Lineal y el Propensity Score como indicadores del tamaño del efecto de un determinado tratamiento en una encuesta compleja. Dicha encuesta indagó sobre diferencias salariales de género en Estados Unidos, controlando por covariables de tipo educativa y laborales. La conclusión más importante de este estudio fue la similitud de los estimadores en cuanto al efecto encontrado y la significancia estadística de los mismos. Por otro lado, Cepeda y colaboradores en su estudio del 2004 (6) realizaron un estudio comparativo entre la regresión logística contra el Propensity Score en búsqueda de diferencias con respecto a términos de sesgo, precisión, probabilidad de cobertura empírica, poder empírico al detectar diferencias estadísticamente significativas y la robustez cuando el número de eventos es bajo con relación con el número de factores de confusión, encontrando que el sesgo disminuyó a medida que el número de eventos por factor de confusión aumentaba. Para el Propensity Score se encontró que el sesgo disminuyó a medida que la fuerza de la asociación entre la exposición y el desenlace aumentaba. Por último, Linden en el 2014 (22) describió las características, supuestos y ventajas de la implementación del MMWS (Ponderación Media Marginal por Estratificación) como estrategia complementaria en el uso eficiente del Propensity Score y destaca a esta metodología como flexible, interesante, de amplia aplicabilidad y como una alternativa importante al momento de evaluar la efectividad en intervenciones en salud. Ante las nuevas metodológicas que permiten la estimación del Propensity Score para tratamientos o exposiciones de tipo categórico nominal u ordinal con el fin de determinar

el efecto de una determinada exposición o intervención (22) y que según la literatura se destacan por el nivel relativamente alto de robustez, a pesar de los problemas de especificación del mismo Propensity Score convencional y la visible ganancia de eficiencia que implica el uso de este método en cuanto a la estimación del efecto (22,23), es necesario determinar bajo qué escenarios estas nuevas herramientas son aplicables y determinar la ganancia que éstas pueden ofrecer con respecto a los métodos lineales que aún siguen siendo aplicado en el contexto epidemiológico con frecuencia.

Aunque se han hecho sugerencias del número de sujetos y condiciones mínimas necesarias para encontrar resultados no sesgados al usar un análisis de regresión lineal múltiple y regresión ordinal (24,25) y para el uso del Propensity Score (6), éstas no consideran la naturaleza del desenlace y el cambio en la relación expuesto-no expuesto 1:1. Adicionalmente, la mayoría de los estudios de simulación no sugieren un tamaño de muestra mínimo necesario para el uso combinado de estas estrategias de control de confusión (análisis doblemente robusto).

Por último, ante lo relativamente reciente de la metodología MMWS, aún no existen evidencia clara en la literatura que demuestre que este método de estimación en comparación por ejemplo con un modelo de regresión lineal múltiple provee mejores ajustes en cuanto al fenómeno de confusión, esto adquiere especial importancia debido al creciente uso del Propensity Score Ordinal basado en MMSW como metodología de control de confusión. En la literatura se han encontrado algunos estudios que aplican estas metodologías (de regresión lineal o Propensity Score (16)) de manera indistinta, sin embargo, dichos estudios no han determinado si estos métodos de ajuste de la confusión son óptimos para este fin. Adicionalmente, no se ha evaluado que condiciones proporcionan un mejor ajuste en términos de la estimación del efecto en un tratamiento con distintos niveles de exposición.

2. Pregunta de investigación

La justificación del problema hace un breve resumen de los diversos estudios que han buscado determinar las mejores condiciones para la aplicación de las metodologías comparadas en este estudio: Propensity score o regresión lineal múltiple. En esta breve descripción se denotan hallazgos importantes y vacíos del conocimiento que es importante resolver en términos de que metodología debe aplicarse ante determinado escenario analítico que presente un estudio de tipo observacional. Es por lo anterior, que surge la siguiente pregunta de investigación:

¿Cuál es la mejor metodología (modelo lineal ordinario o Propensity Score combinado con MMWS) para controlar el fenómeno de la confusión en estudios observacionales analíticos, teniendo en cuenta factores como el tamaño de muestra, la relación expuesto-no expuesto y número de variables a incluir?

3. Marco teórico

3.1 Conceptos estadísticos

3.1.1 Modelos Lineales

Los modelos lineales generalizados fueron formulados por John Nelder y Robert Wedderburn con el fin de unificar varios modelos estadísticos, incluyendo la regresión lineal, regresión logística y regresión de Poisson, entre otras bajo un solo marco teórico (26). Esto les permitió desarrollar un algoritmo general para la estimación de máxima verosimilitud en todos estos modelos, teniendo como base la teoría establecida del modelo de regresión lineal para la estimación, validación y uso de modelos (27,28). Las técnicas de regresión múltiple son más adecuadas para afrontar problemas relacionados con la implementación de otras técnicas que buscan resolver el efecto de la confusión, ya que permite examinar los efectos de todas las variables de exposición simultáneamente y de forma recíproca (9), permitiendo ajustar el efecto de una variable de interés tanto por variables discretas como continua, determinando el peso o impacto de cada una de las variables consideradas independientes en dicho desenlace (9,23,24).

Como se mencionaba anteriormente, el análisis de regresión es una de las técnicas de uso más frecuente para analizar datos multifactoriales (29), dado que es un método conceptualmente simple para investigar las relaciones funcionales entre variables (30). Para un modelo de regresión lineal se tiene que la relación de una variable respuesta con un conjunto de covariables está dada por (29):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Donde los distintos β se llaman coeficientes de regresión y miden la contribución de la variable X_i en la respuesta Y y donde se supone que el término del error ε del modelo tiene $E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2$, adicionalmente los errores no deben estar correlacionados y deben

seguir una distribución normal. (27,29,30). Estas condiciones en especial esta última permiten comprobar las hipótesis de los parámetros del modelo y estimar los intervalos de confianza de los mismos. Otro concepto importante en el análisis de regresión íntimamente relacionado con el concepto de error es el concepto de residual, que se define como la diferencia entre el valor observado y el valor ajustado dado por el modelo, indicando la desviación existente entre los datos observados y el ajuste por el modelo de regresión (29). Matemáticamente estos se define como (27,29):

$$e_i = y_i - \hat{y}_i$$

Los residuales juegan un papel fundamental en el diagnóstico de violaciones de los supuestos de un modelo de regresión (29,30). Dichos supuestos a verificar en un modelo de regresión son (24,29):

- Linealidad: Este supuesto asume que la relación entre las variables independientes con la variable dependiente debe ser lineal. Un coeficiente de correlación significativo permite la verificación de este modelo. Adicionalmente se sugiere que este coeficiente de correlación sea superior a 0.5 pero es un punto de corte arbitrario (30).
- Independencia de los errores: Este supuesto asume que los residuos no están autocorrelacionados, en otras palabras, son independientes. La presencia de autocorrelación produce un problema de subestimación en los intervalos de confianza y las conclusiones sobre los coeficientes del modelo pueden ser erróneas (29). Hay pruebas estadísticas que permiten chequear este supuesto, entre ellas está la prueba de Durbin-Watson.
- Varianza constante: Este supuesto asume que la varianza no varía en los diferentes niveles de un factor, es decir, entre diferentes grupos. Con respecto a los modelos de regresión existe heterogeneidad (varianza no constante) si los residuos tienden a aumentar o disminuir con los valores de la variable de predicción (30). El incumplimiento de este supuesto afecta negativamente la precisión de los intervalos de confianza de los estimadores del modelo. La prueba de Levene permite chequear este supuesto (29) y el diagnóstico se complementa con análisis gráficos de los valores estimados contra las variables independientes del modelo (30).

- Normalidad en los errores: Este supuesto asume que los errores del modelo deben seguir una distribución normal. El incumplimiento de este supuesto trae consigo problemas en el uso de las pruebas de hipótesis de los parámetros del modelo y problemas de imprecisión en los intervalos de confianza creados para el modelo (30). Una prueba de Shapiro-Wilk, permite evaluar este supuesto, en conjunto con herramientas gráficas (10,24,29).
- Los errores tienen una media de cero (24,30).
- No colinealidad: Este supuesto asume que no existen relaciones lineales exactas entre las variables independientes. La violación de este supuesto produce estimadores “demasiado grandes” y aumenta la variabilidad de los estimadores del modelo (29). El VIF (factor de inflación de varianza) es un indicador que permite corroborar este supuesto (30).
- Las variables predictoras son constantes conocidas y no aleatorias, medidas sin error (29,30).

En un modelo lineal generalizado, lo que se busca es mayor flexibilidad de los supuestos que rigen a la regresión lineal convencional (31). Uno de los supuestos sobre la distribución de la variable respuesta, ya que el supuesto de normalidad se deja de lado y se suponen más distribuciones para el desenlace de interés. Formalmente hablando se tiene que el primer supuesto de un modelo lineal generalizado (GLM) es (27):

- La distribución de la variable respuesta y_i pertenece a la familia exponencial y depende de un solo parámetro θ_i , es decir (27):

$$f(y_i; \theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]$$

Donde $b_i(\theta_i)$ es conocido como *parámetro natural*, que es una función que depende de μ_i y los demás componentes de la función representan parámetros de escala y de dispersión de cada distribución. Dicho de otra forma, un modelo lineal generalizado plantea una relación funcional, basada en una función conocida como función de enlace g de la forma:

$$g(\mu_i) = \underbrace{x_i'}_{\beta}$$

Esta función de enlace es uno de los tres componentes de un modelo lineal generalizado, dichos componentes se definen a continuación (27,32):

- Una componente aleatoria: Variables respuestas Y_1, Y_2, \dots, Y_n con la misma distribución perteneciente a la familia exponencial,
- Una componente sistemática $X\beta$, compuesta por un conjunto de parámetros $\beta = [\beta_0, \beta_1, \dots, \beta_p]$ y un conjunto $X = [x_1, x_2, \dots, x_k]$ de variables explicativas.
- Un componente funcional dado por la función de enlace monótona g que proporciona una relación funcional entre la componente sistemática y la esperanza de la respuesta en la componente aleatoria (27,29).

$$g(\mu_i) = x_i' \beta \quad \text{donde} \quad \mu_i = E(Y_i)$$

3.1.2 Regresión logística

La regresión logística es un tipo de modelación enmarcada en la teoría de Modelos Lineales Generalizados, que busca estimar la probabilidad de un suceso dado un conjunto de variables independientes (32). El objetivo en este caso, es determinar el grado de asociación entre una variable que mide la presencia y ausencia de una característica en particular y un grupo de variables que pueden afectar la ocurrencia de dicho evento (27,32). Para este tipo de regresión generalizada la función de enlace se conoce como *logit* que se define como (25):

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \ln(p) - \ln(q)$$

Donde:

$p = \text{probabilidad de ocurrencia de un evento}$

$q = 1 - p$

$\ln(x)$ Es la función logaritmo en base e .

Para este modelo se asume que la variable respuesta se distribuye de manera binomial es decir $Y_i \sim B(p_i, n_i)$, donde el número de ensayos es conocido y se busca estimar la probabilidad de ocurrencia del evento p_i . Usando la función *logit* es posible modelar estas probabilidades dado un conjunto de variables de la forma:

$$\text{logit}[P(Y = 1)] = \text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$$

Donde cada probabilidad ajustada es de la forma (27):

$$p_i = \frac{1}{1 + e^{\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}$$

Los modelos de regresión para resultados binarios basados en funciones *logit* son la base a partir de la cual pueden derivarse modelos más complejos para datos ordinales y nominales (24,30–32). Los modelos de regresión ordinal y nominal son equivalentes a la estimación simultánea de una serie de resultados binarios (32). Bajo este orden de ideas, se establece un modelo que compute las probabilidades de ocurrencia de una variable respuesta con K categorías, a través de una forma funcional como se muestra a continuación (32):

$$P(Y = j | \mathbf{x}) = P(\tau_j \leq Y^* < \tau_{j+1} | \mathbf{x})$$

O en términos de la función *logit* (26,27):

$$\begin{aligned} \text{logit}[P(Y \leq j)] &= \log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) \\ &= \log\left(\frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \dots + \pi_J}\right) \end{aligned}$$

Estas funciones se conocen como *logits* acumulados que muestra la probabilidad de estar en la j -ésima categoría (27). Dado lo anterior, para una variable predictora X , el modelo *logit* puede escribirse de la forma:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta X, \quad j = 1, 2, \dots, J - 1$$

Donde el parámetro β que describe el efecto de X sobre el *logit* de la respuesta en la categoría j o inferior. En esta fórmula, β no tiene subíndice j , es decir, el modelo asume un efecto idéntico de X para todos los $J-1$ colapsos de la respuesta en resultados binarios. Si existe un buen ajuste del modelo, se requiere un solo parámetro en lugar de $J-1$ parámetros para describir el efecto de X (27), dado que los efectos de las variables predictoras son iguales para cualquiera de las categorías (32). Una prueba de Brant permite comprobar este supuesto (31) y ante la violación de este supuesto llamado “supuesto de regresión

paralela” o “de odds proporcionales” pueden usarse modelos logísticos generalizados que tienen en cuenta el cambio diferencias en los chances de oportunidad de la variable respuesta (32).

3.1.3 Función Kernel

Dada la muestra de n observaciones reales X_1, \dots, X_n se define la estimación tipo kernel K como (33):

$$\hat{f}(x) = \frac{1}{n h} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

Donde $K(x)$ se define como la función Kernel y su función es estimar una función de densidad que tenga un modelo desconocido o no paramétrico, con base en los valores muestrales (34). El valor h se conoce como ancho de banda y se encarga de ponderar la función kernel con base en el nivel de influencia dado a cada valor observado. Existe una amplia gama de funciones Kernel a usar (tabla 3-1), aunque cualquiera es válida es importante realizar análisis de sensibilidad y validación cruzada (35,36) para evaluar posibles valores de este parámetro (37).

Tabla 3-1: Principales funciones Kernel (33).

Núcleo	$K(t)$	Rango
Epanechnikov	$\frac{3}{4}(1 - t^2)$	$ t < 1$
Gaussaiano	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$	$ t < \infty$
Triangular	$ 1 - t $	$ t < 1$
Rectangular	$\frac{1}{2}$	$ t < 1$
Biweight	$\frac{15}{16}(1 - t^2)^2$	$ t < 1$
Triweight	$\frac{35}{32}(1 - t^2)^3$	$ t < 1$
Arco coseno	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}\right) t$	$ t < 1$

Las funciones Kernel tienen múltiples aplicaciones, desde estimación no paramétrica de funciones de densidad, hasta el uso de este en el emparejamiento en un Propensity Score basado en la distancia entre los individuos de dos grupos de interés (expuestos - no expuestos, intervenidos - no intervenidos) con el fin de estimar las diferencias existentes en dichos grupos con respecto a una variable de interés (1,38).

3.1.4 Propensity Score Matching

Es una herramienta estadística muy utilizada en medicina, que pretende determinar el impacto de una intervención o exposición en una variable de interés por cuenta de covariables que permiten determinar la probabilidad de recibir dicha intervención (39). El Propensity Score busca construir una función que estime la probabilidad individual de recibir un tratamiento condicionado a un conjunto de covariables, esta probabilidad está definida como $P(X) = P(D = 1|X_1, X_2, X_3, \dots, X_n)$, con esto se busca reducir el sesgo debido al efecto de confusión, dado que busca construir grupos homogéneos y así identificar posibles diferencias en la variable de interés de una manera no sesgada. (1,23)

Esta comparación mencionada anteriormente entre los grupos tratamiento en un estudio observacional puede verse afectada por factores que estén asociados con la intervención y la variable respuesta (40), ahora bien, una estimación de dicho efecto será insesgada si (28):

- Existe independencia de efectos a nivel individual, es decir, el conjunto de variables evaluadas se distribuye homogéneamente entre tratamientos y controles.
- Existe una asignación de intervenciones “no confundida”, donde la aleatorización y posterior asignación en tratamientos y controles es independiente de la variable respuesta.
- Se cumple la “Asunción de Valor de Tratamiento de Unidad Estable” (SUTVA por sus siglas en inglés) (41), este supuesto asume que el resultado observado en el sujeto i con un tratamiento asignado t es igual, sin importar el mecanismo de asignación de dicho tratamiento y de los demás sujetos aleatorizados.
- No existe sesgo de selección en las covariables evaluadas y no hay variables que no hayan sido medidas que puedan confundir el efecto de la variable desenlace.

Cumpléndose estas características es posible encontrar diversos estimadores del efecto en cuestión, entre los estimadores más conocidos están:

- Estimador de efecto promedio (Average Treatment Effect -ATE):

$$ATE = \tau = E(Y_1 | X, W = 1) - E(Y_0 | X, W = 0)$$

Donde $Y_{i(1)}$ representa a un individuo i expuesto a una intervención, $Y_{i(0)}$ representa un individuo i no expuesto a una intervención, X representa el conjunto de covariables usadas para calcular dicho efecto y W indica si el sujeto es tratamiento (1) o control (0).

Visto desde el punto de vista estadístico, el ATE es el estimador usado en un estudio experimental o cuando se considera que no hay sesgo de selección, en estos casos se usan prueba T Student, ANOVA o sus pruebas no paramétricas (41).

- Estimador de Efecto promedio en los tratados (Average Treatment effect on the Treated-ATT) (32,33):

$$ATT = \tau = E(Y_1 - Y_0 | X, W = 1)$$

En este estimador se obtiene la diferencia entre el resultado del paciente tratado y el control dado un conjunto de covariables y dicho efecto se obtiene promediando todas estas diferencias, dando así una estimación del efecto del tratamiento en los pacientes con tratamiento.

- Estimador de Efecto promedio en los no tratados (Average Treatment effect for the Untreated -ATU) (32,33):

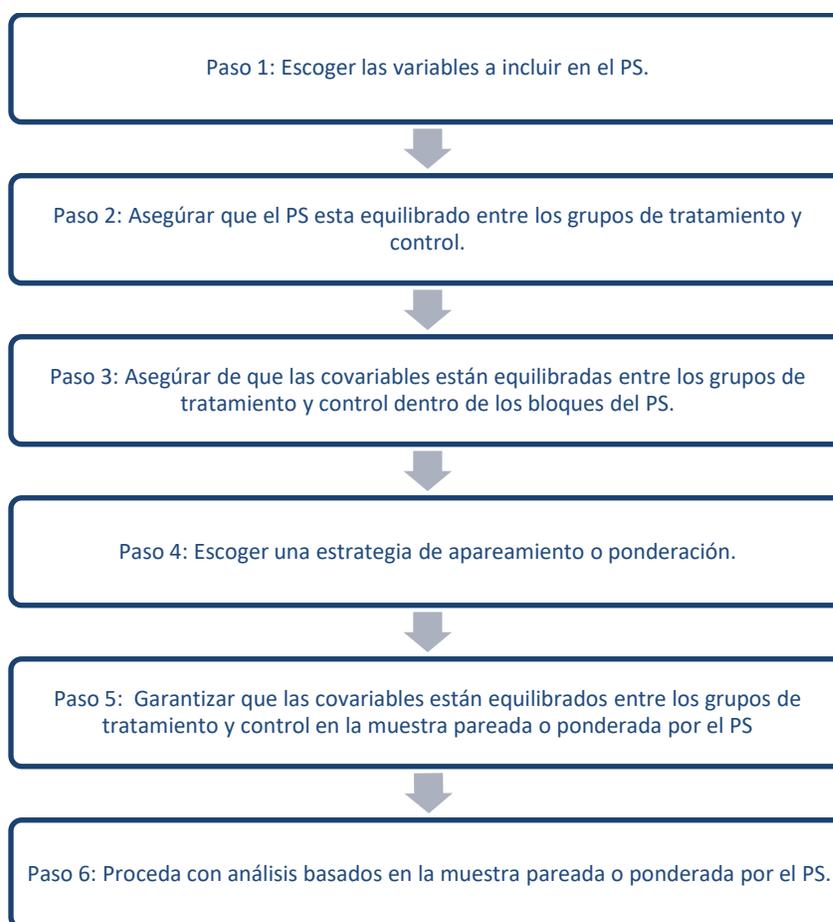
$$ATU = \tau = E(Y_1 - Y_0 | X, W = 0)$$

Similar al estimador anterior, busca determinar un diferencial estimado entre aquellos pacientes sin tratamiento con los pacientes intervenidos, la idea de este estimador es resolver la duda sobre el efecto de la intervención, si se tratara con ella a los controles.

3.1.5 Pasos para la estimación de un Propensity Score

Con el fin de estimar de manera óptima este efecto y evaluar su significancia en la literatura se proponen varias metodologías para la utilización del Propensity score como herramienta de comparación, pero en general se siguen los pasos descritos en la figura 3-1 (41,44). Nótese que los pasos 1 a 5 son secuenciales y puede que tenga que ejecutarse varias veces, ya que en algunas ocasiones puede presentarse desbalance en las variables incluidas en el Propensity Score, problemas de soporte común entre los sujetos tratados y no tratados o heterogeneidad en alguno de los bloques construidos.

Figura 3-1: Pasos para la implementación para un Propensity Score (44)



- **Selección de variables**

En este paso se deben escoger las variables que se consideren pertinentes para la construcción del Propensity Score. Ante la duda de cuales variables deben incluirse y bajo qué criterio se hace la selección se ha sugerido que se deben incluir aquellas variables

que se encuentren asociadas con el desenlace y con la intervención o tratamiento (1,28,43). Adicionalmente debe tenerse en cuenta el área de acción donde se esté aplicando esta metodología ya que pueden omitirse variables importantes, generando así una fuente de sesgo en el análisis (7). Bajo este marco de selección se percibe la existencia de 3 tipos de covariables (41,45):

- a. Las covariables que se asocian con el tratamiento y con la variable respuesta.
- b. Las covariables asociadas con la variable respuesta y no con el tratamiento.
- c. Las covariables relacionadas con el tratamiento y no con la variable respuesta.

En el Propensity Score deben incluirse las variables mencionadas en los apartados a. y b. ya que las primeras son fuentes de confusión y las segundas contribuyen a una disminución de la variabilidad. Debe tenerse en cuenta que la inclusión de muchas variables puede causar inflación de la varianza y posiblemente no contribuyan con la reducción del sesgo (38,41,45).

▪ **Equilibrio entre tratamientos y controles**

Un concepto importante en este aspecto es el “soporte común” que se define como la región de solapamiento de las probabilidades estimadas de asignación. Se deduce de la definición que tanto casos como controles tenga una superposición de dichas distribuciones con el fin de garantizar una comparación equitativa entre ambos grupos (28). La violación de este supuesto puede llevar a la inclusión de sesgo en el análisis, ya que se puede estar comparando individuos que de entrada no son comparables (44).

▪ **Equilibrio entre las covariables**

En la construcción del PS se lleva a cabo una estratificación con el fin de comparar de manera óptima un subgrupo de tratamientos contra controles, en caso de no presentarse este balanceo se pueden dar estimaciones sesgadas como en el caso anterior (43). De presentarse esta situación debe considerarse la inclusión de interacción entre las variables seleccionadas, la inclusión de nuevos factores o la eliminación de covariables del modelo (41,45).

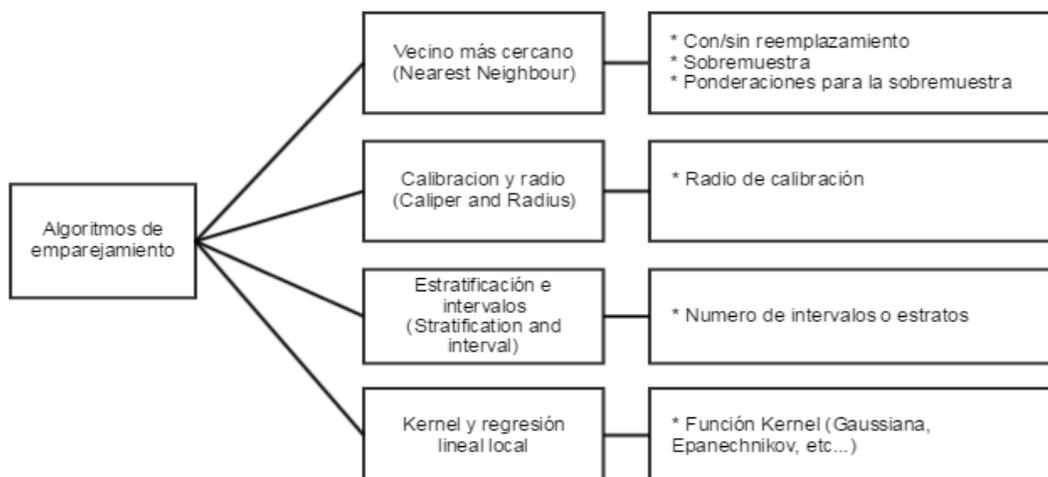
▪ **Estrategia de emparejamiento**

Con el “Matching” o emparejamiento se busca comparar un sujeto tratamiento con uno o varios controles dado que tienen un Propensity Score similar (46). Un problema que puede presentarse en el emparejamiento es una reducción importante del tamaño de la muestra

(9,43), dado que un tratamiento podría no encontrar un control apropiado para emparejar, adicionalmente es posible que, si no hay suficiente superposición entre los grupos, se pueden comparar los peores controles contra los mejores tratamientos (28).

Existen muchos tipos de emparejamiento (12,43), algunos de los métodos clásicos más usados se describen a continuación:

Figura 3-2: Tipos de emparejamiento (12,43)



- *Vecino más cercano:* en este método el emparejamiento se realiza entre los individuos tratamiento y control con el Propensity Score más cercano, puede darse que un paciente control puede hacer “matching” con más de un tratamiento bajo esta metodología (43). Una de las dificultades de este método es que las correspondencias entre tratados y controles pueden ser deficientes en el sentido que un control podría encontrar como vecino más cercano a un tratamiento con un Propensity Score totalmente opuesto (38,45).
- *Emparejamiento por radio:* para este caso, se define “a priori” una vecindad demarcada por un radio en función de la desviación estándar del Odds del Propensity Score definido como un quinto de la desviación estándar (1) del logit de los valores del PS $\left(\text{logit } P(Z) = \ln\left(\frac{P(Z)}{1-P(Z)}\right) \right)$ (38,45). Un inconveniente de este emparejamiento es que con radios muy pequeños un paciente tratamiento no encuentre su par control, sin embargo, de encontrar un control correspondiente su apareamiento será óptimo. Otro inconveniente con este método es el aumento de varianza dado que puede haber pocos controles (28). Sin embargo, Dehejia y

Wahba (47) sugieren que esta variante de al usar no solo el vecino más cercano dentro de cada radio, sino todos los miembros de comparación dentro de la radio provee una ventaja ya que usa solo tantas unidades de comparación como están disponibles dentro del radio definido y por lo tanto, permite el uso de unidades adicionales (menos) cuando las coincidencias buenas no están (no) disponibles, asimismo este tipo de emparejamiento reduce el número de coincidencias erróneas entre expuestos y no expuestos aumentando la calidad de la coincidencia, permitiendo una minimización del Error Cuadrático Medio y una reducción superior al 90% en el sesgo (48,49).

- *Emparejamiento/ponderación por método Kernel* (34): el “matching” en este caso se da entre un sujeto tratado con un promedio ponderado de los individuos control, donde la ponderación w_{ij} viene dada por el inverso de la distancia entre los PS de los grupos a comparar, donde (37):

$$w_{ij} \propto K\left(\frac{p_i - p_j}{h}\right)$$

Así las cosas w_{ij} es un promedio ponderado de los resultados de (posiblemente todas) las unidades no tratadas donde el peso otorgado a la unidad j no tratada es proporcional a la cercanía de los observables de i y j y h es el ancho de banda correspondiente (50,51).

Cabe resaltar que como se mencionaba anteriormente existen más métodos de emparejamiento (ponderación por probabilidades inversas de tratamiento IPTW por sus siglas en inglés, emparejamiento por distancias de Mahalanobis, entre otros) (43) y que no hay un método superior a los demás y que la elección de un método particular depende del contexto del problema.

- **Equilibrio entre las covariables después de la estimación del Propensity Score**

Con el fin de evaluar este equilibrio después de construido el Propensity Score, puede usarse diferencias estandarizadas, gráficos estadísticos y evaluación de estadísticas descriptivas para tal fin (44).

Para esto es necesario comparar una vez ejecutado el Propensity Score, los promedios o medianas de las covariables entre tratados y no tratados antes y después del emparejamiento y hacer una inspección visual del cambio en el porcentaje de sesgo estandarizado en todas de las covariables incluidas en el PS (28). Otra fuente de

verificación de este supuesto se basa en el uso de gráficos especialmente los “Q-Q plots”, ya que estos permiten comparar las covariables continuas entre tratamientos y controles y observar posibles discrepancias entre las distribuciones de los grupos (52). Por último, se puede recurrir a una estimación de la densidad vía kernel de las variables continuas con el fin de comparar empíricamente la igualdad de las distribuciones de las variables para el grupo intervención y de control (1,44).

- **Análisis comparativo y de sensibilidad con base en el Propensity Score**

Por último, después de comprobar los anteriores supuestos y garantizar que las covariables incluidas en el análisis están balanceadas, se procede a estimar el efecto del tratamiento con uno de los estimadores vistos anteriormente (Estimador de efecto promedio, Estimador de Efecto promedio en los tratados, Estimador de Efecto promedio en los no tratados), revisando que antes se haya eliminado las observaciones de que están fuera del rango de soporte común, elegido el efecto del tratamiento de interés y calculado el error estándar correcto para el Propensity Score (53). Debe evitarse una mala especificación del PS, ya que esto puede llevar a conclusiones erróneas. Adicionalmente con el fin de mejorar la estimación del efecto de tratamiento y su correspondiente error estándar puede recurrirse a herramientas como el bootstrap (35,36) con el fin de obtener estimaciones más precisas en cuanto al tamaño del efecto y su correspondiente error estándar (54).

Por otro lado, en cuanto al análisis de sensibilidad es recomendable el uso del test de Cochran–Mantel–Haenszel (44) prueba que se usa ampliamente para medir la fuerza de la asociación entre una exposición y una enfermedad o respuesta, después de estratificar las covariables observadas (55), con el fin de evaluar si existen covariables observadas que estén influenciando la variable desenlace (56).

3.1.6 Método de Ponderación Media Marginal Basado En Un Propensity Score a través De estratificación (MMWS)

Las intervenciones que incluyen diversos grados de tratamiento de nivel ordinal también son frecuentes en los estudios epidemiológicos; los estudios de medicamentos en los que se comparan las diferentes dosis de un medicamento son un ejemplo (22). En estos casos, se busca evaluar si existe una relación positiva (o inversa) entre los niveles de tratamiento

o exposición observados y el resultado de interés (22,57). En estudios donde el grupo de comparación o el grupo de control que no recibe esta intervención o el grupo que recibe el tratamiento más bajo las técnicas de ponderación pueden servir para lograr un equilibrio en las características basales entre todos los niveles de tratamiento (22,57,58).

El método de ponderación media marginal basado en un Propensity Score a través de estratificación (MMWS por sus siglas en inglés) es un mecanismo que busca eliminar el sesgo de selección asociado con un gran número de covariables al equiparar la composición de un grupo sin tratamiento con varios grupos de tratamiento (58). El método MMWS además de usarse en tratamientos binarios, es útil en evaluaciones donde se miden distintos niveles de exposición o tratamiento (ordinal) o varias exposiciones o intervenciones (nominal), adicionalmente, otra de las ventajas de este método de ponderación no paramétrico es que según la literatura muestra un nivel relativamente alto de robustez a pesar de especificaciones erróneas en el modelo funcional y también una notable eficiencia y precisión en los estimadores que otorga en comparación con otras técnicas (57,59).

Al evaluar si los efectos del tratamiento difieren entre las subpoblaciones definidas por las características individuales o por los niveles del tratamiento, se pueden asignar pesos dentro de cada subpoblación con el fin de aproximar un diseño de bloques al azar (22). El método también puede usarse para evaluar si el efecto de un tratamiento inicial se amplifica o se debilita mediante un tratamiento subsiguiente o para identificar una secuencia de tratamiento óptima mediante la aproximación de un experimento aleatorio secuencial (57,58).

MMWS usa un procedimiento no paramétrico que calcula estos pesos sobre la base de la estratificación del puntaje de propensión, estas ponderaciones quedan definidas en función de la probabilidad de pertenecer a cada una de las exposiciones dada las covariables que predicen esta intervención y el número de individuos de cada exposición y son de la forma:

$$MMW = \frac{p(\theta_z)}{p(\theta_z|Z = z)} = \frac{n_s}{n_{z,s}} \times p(Z = z)$$

Donde n_s es el número total de individuos en un estrato s , $p(Z = z)$ es la probabilidad de asignación al grupo de tratamiento z , y $n_{z,s}$ es el número total de individuos en los estratos que en realidad fueron asignados a tratamiento z (59). Estas ponderaciones son usadas

como pesos de muestreo en el modelo de asociación que busca determinar el efecto los tratamientos/exposiciones de interés (22,57).

Bajo el supuesto de ignorabilidad que indica que la elección de asignación de un sujeto a un grupo de control o un grupo de tratamiento puede suponerse efectivamente como aleatoria cuando está condicionada a las características observables de los objetos de estudio, y que los datos faltantes pueden tratarse como si ocurrieran efectivamente al azar (60), los datos ponderados deben aproximarse a un experimento aleatorio. Los pesos generados por esta metodología pueden combinarse con otras metodologías para disminuir adicionalmente el sesgo residual y/o aumentar la precisión (41,45).

El procedimiento analítico general de ajuste de un MMWS implica ocho pasos principales que se aplican de manera similar en tratamientos multivalorados de tipo nominal u ordinal (57).

1. Recolección de un gran conjunto de covariables de pretratamiento.
2. Estimación del Propensity Score.
3. Identificación del soporte común.
4. Estratificación de la muestra analítica.
5. Cálculo de los pesos no paramétricos.
6. Comprobación del equilibrio en la composición de covariables entre los grupos de tratamiento en la muestra ponderada.
7. Análisis de un modelo de resultados ponderado para estimar los efectos del tratamiento.
8. Realización de un análisis de sensibilidad.

Como se puede observar estos pasos son similares a los pasos que deben realizarse para la estimación de un Propensity Score.

3.1.7 Estimadores doblemente robustos

Con el fin de mitigar el sesgo de selección que surge de la falta de respuesta y el fenómeno del desgaste, la asignación del tratamiento no aleatorio en estudios observacionales y el incumplimiento del proceso de aleatorización en Ensayos Clínicos Aleatorizados (61) se han diseñado nuevos métodos de estimación conocidos como estimadores doblemente robustos (62).

Estos métodos de estimación hacen uso de la teoría de estimación ponderada para datos incompletos (61) y buscan mitigar el fenómeno de la pérdida de información, modelando las probabilidades de censura dadas las covariables e incorporar dichos puntajes en una estimación ponderada o estratificada con el fin de mejorar las estimación en una intervención. Esta metodología, sin embargo, fue implementada en los modelos de Propensity Score con el fin de robustecer las estimaciones dadas por estos puntajes de propensión (22,63). Emsley y colaboradores (64) proponen un algoritmo para la implementación de este tipo de estimadores:

1. Ajustar un modelo de regresión logística (o probit) para el tratamiento condicional en la línea de base variables (variables dependientes del tiempo pueden incluirse si es necesario para análisis). El valor predicho de esta regresión da como resultado el puntaje de propensión estimada p_i
2. Ajustar un modelo de regresión para la variable respuesta para los individuos que presenten el tratamiento o la intervención de interés (es decir $j = 1$) y obtener la predicción de estos valores.
3. Ajustar un modelo de regresión para la variable respuesta para los individuos sin el tratamiento o la intervención de interés (es decir $j = 0$) y obtener la predicción de estos valores
4. Obtener los estimadores doblemente robustos definidos como (62):

$$\hat{\mu}_{OLS} = \frac{1}{n} \sum \hat{m}_i$$

Donde:

- $\hat{m}_i = x_i^T \hat{\beta}$
- $\hat{\beta} = (\sum_j t_j (\hat{\pi}_j)^{-1} x_j x_j^T)^{-1} (\sum_j t_j (\hat{\pi}_j)^{-1} x_j y_j)$

De esta forma se establecen estimadores que tiene en cuenta la propensión a estar expuesto o no a una condición de interés y bajo este concepto las estimaciones son más precisas (22). Nótese que bajo la teoría que envuelve a los modelos de regresión se está estimando un modelo de mínimos cuadrados ponderados (30), donde los pesos quedan determinados por el Propensity Score (58,62,63).

3.1.8 Métodos de comparación de modelos

Las metodologías de comparación utilizadas en este estudio se basan en el error de los modelos en cuestión: regresión lineal múltiple y Propensity Score combinado como MMWS. El error de estimación de un modelo se define como la diferencia existente entre el parámetro de interés y el valor estimado (27,30).

Con el fin de comparar las metodologías establecidas en este estudio se definen los siguientes índices:

- a) **Error Porcentual Absoluto Medio (MAPE):** Es un indicador de desempeño que mide el tamaño del error en términos porcentuales. Se define como:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\theta_i - \hat{\theta}_i}{\theta_i} \right|$$

- b) **Diferencia promedio estandarizada:** Se define como la diferencia porcentual entre un parámetro con su valor estimado

$$\% SMD = \sum_{i=1}^k \left(\frac{\hat{\theta}_i - \theta_i}{\theta_i} \right)$$

- c) **Error Cuadrático Medio:** Matemáticamente se define como la suma residual de cuadrados, dividida por el número de grados de libertad. El Error Cuadrático Medio en un estimador es un indicador determinado por el promedio de los errores al cuadrado, indicando la diferencia entre un parámetro y su valor estimado.

$$MSE = \sum_{i=1}^n \frac{(e_i)^2}{n} = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}$$

4. Objetivos

4.1 Objetivo general

Determinar que metodología estadística es más apropiada entre la estimación de un modelo de regresión lineal convencional y la estimación del tamaño del efecto por Propensity Score Ordinal usando Ponderación Media Marginal por Estratificación (MMWS) para controlar la confusión bajo diversos escenarios de tamaño de muestra, relación expuesto-no expuesto y un número de predictores fijos.

4.2 Objetivos específicos

- Comparar el desempeño de las metodologías: ajuste por modelos lineales o Propensity Score Ordinal, para los escenarios de tamaño de muestra y relación expuesto-no expuestos establecidos, a través del Error porcentual absoluto medio (MAPE), Diferencia promedio estandarizada (SMD) y el Cuadrado Medio del Error (MSE).
- Sugerir las condiciones mínimas necesarias en cuanto a tamaño de muestra para el uso de estimadores doblemente robustos (definiendo esto estimadores como la combinación de las metodologías de Propensity score y modelo lineal) en un estudio epidemiológico de tipo analítico observacional, con el fin de controlar de manera eficiente el fenómeno de la confusión.

- Evaluar el impacto del uso de un mayor número de no expuestos por expuestos en el control de la confusión en estudios observacionales.

5. Metodología

5.1 Tipo de investigación y diseño

Estudio de simulación que se llevará a cabo entre los meses de octubre del 2017 y enero de 2018. Se plantea este tipo de estudio debido a que la literatura publicada sobre este tema es mayoritariamente teórica y no se ha establecido un análisis comparativo que permita determinar que metodología es más eficiente en el control de la confusión.

Se justifica un estudio de simulación, ya que se pretende realizar un análisis de sensibilidad donde se determine el impacto de distintos tamaños de muestra y relación expuesto-no expuesto en la estimación del efecto.

5.2 Hipótesis

Las hipótesis planteadas en este proyecto son:

H₀: Bajo condiciones controladas de tamaño de muestra, relación expuesto-no expuesto, el estimador del efecto dado por un modelo lineal generalizado es igual al estimador dado por un análisis de Propensity Score Ordinal basado en MMWS.

$$H_0: \theta_{GLM} = \theta_{PSMMWS}$$

H_a: El estimador del efecto dado por un modelo lineal generalizado es diferente al estimador dado por un análisis de Propensity Score Ordinal basado en MMWS.

$$H_1: \theta_{GLM} \neq \theta_{PSMMWS}$$

5.3 Análisis comparativo y de sensibilidad

5.3.1 Relación expuestos-no expuestos

Con el fin de determinar el efecto que tiene un número de casos por control se tiene presupuestado realizar un análisis de simulación con relación expuesto-no expuesto de 1:1, 1:2 y 1:4, dado que son las relaciones más frecuentes en los estudios epidemiológicos (9,28).

5.3.2 Escenarios de tamaños de muestra

Se realizarán simulaciones del tamaño de muestra bajo diferentes escenarios, considerando un tamaño de muestra mínimo de 100 sujetos hasta un máximo de 2000 con el fin de evaluar la exactitud de los estimadores para cada uno de los métodos evaluados, estos escenarios abarcan los tamaños de muestra que generalmente se muestran en los estudios observacionales en la actualidad. Los escenarios se muestran a continuación.

Tabla 5-1: Escenarios de tamaño de muestra establecidos para el análisis

Escenario de tamaño de muestra										
	Tamaño de muestra por grupo	relación 1:1			relación 1:2			relación 1:4		
		No expuesto	Expuesto nivel 1	Expuesto nivel 2	No expuesto	Expuesto nivel 1	Expuesto nivel 2	No expuesto	Expuesto nivel 1	Expuesto nivel 2
Escenario 1	100	100	100	100	200	100	100	400	100	100
Escenario 2	200	200	200	200	400	200	200	800	200	200
Escenario 3	500	500	500	500	1000	500	500	2000	500	500
Escenario 4	750	750	750	750	1500	750	750	3000	750	750
Escenario 5	1000	1000	1000	1000	2000	1000	1000	4000	1000	1000
Escenario 6	2000	2000	2000	2000	4000	2000	2000	8000	2000	2000

5.3.3 Número de variables a incluir

El número de variables a incluir en el análisis será inicialmente de 6 variables, se incluirán tanto variables de tipo cualitativas y cuantitativa con el fin de acercarse a un escenario real. Para las variables cualitativas se definirán variables categóricas a través de un proceso de segmentación de variables continuas, mientras que las variables continuas serán de tipo normal. Adicionalmente, con fines de evaluar el efecto de la inclusión de más variables en el ajuste de los modelos comparados se simularán consecutivamente 4 variables adicionales para comparar el efecto que estas tienen en la estimación final del efecto de interés.

5.3.4 Análisis de la información

Se realizará un análisis de simulación, con un total de 1000 simulaciones para cada uno de los escenarios evaluados. La información será procesada en el software estadístico R-3.4.2.

Los análisis se realizarán bajo varios escenarios de tamaño de muestra y relación expuesto-no expuesto con el fin de realizar un análisis de sensibilidad con base en estos aspectos. Adicionalmente con el fin de determinar la metodología más eficiente para el control de la confusión (Propensity Score Ordinal basado en MMSW o regresión lineal múltiple) se calcularán algunos índices de precisión, entre ellos el error porcentual absoluto medio (MAPE en inglés), el error cuadrático medio (MSE en inglés) y diferencia promedio estandarizada (SMD en inglés) para cada una de las muestras simuladas.

Los resultados obtenidos serán comparados con los obtenidos para cada una de las metodologías mencionadas.

5.4 Proceso de simulación del estudio

En este estudio de simulación, se compara el estimador MMWS con un estimador basado en una regresión lineal múltiple, en términos de reducción de sesgo. Adicionalmente, este estudio busca determinar las propiedades de los estimadores doblemente robustos, resultado de combinar estas metodologías, realizando un análisis de regresión ponderada por los pesos estimados a través de la metodología MMWS, método conocido como estimadores DR-MMWS (22,59,65).

El diseño de simulación usa los algoritmos descritos por Hong (58) y se basa en dos funciones definidas por Simmering (66), las cuales fueron modificadas para realizar la estimación de los pesos del MMWS posterior a la estimación del Propensity Score basado en una regresión logística ordinal. En este aspecto es importante indicar que se asume que el supuesto de riesgo proporcionales se cumple para el uso de una regresión logística ordinal (32). Se usó el método de emparejamiento por vecino más cercano a través del método de calibración por radio, en este método los individuos del grupo de comparación se emparejan con un compañero coincidente para un individuo tratado que se encuentra dentro del calibre (llamado "rango de propensión") y es el más cercano en términos de puntaje de propensión. El uso de este método de emparejamiento se debe a que este "evita las coincidencias erróneas entre expuestos y no expuestos, aumentando la calidad de la coincidencia (48)" y además produce una minimización del Error Cuadrático Medio (MSE) y una reducción importante del sesgo (49). En cada escenario, se extraen 2000 réplicas del proceso de generación de datos que se describe a continuación y se repiten para tamaños de muestra entre 100 a 2000, con una relación de expuestos: no expuestos entre 1:1 hasta 1:4. Con fines de comparación de los modelos en cuestión, se registran la estimación del efecto del tratamiento, Errores Estándar y valores p empíricos para cada uno de ellos. El sesgo (la diferencia entre el efecto simulado y el efecto verdadero), el error absoluto medio (MAPE) y el Cuadrado Medio del Error (MSE), se calculan en todas las muestras. Los valores más bajos para todas las medidas indican una mejor reducción del sesgo y una mayor precisión en las estimaciones presentadas.

5.5 Proceso de generación de datos para el modelo de tratamiento

Para todos los procesos de generación de datos, se simuló un conjunto de covariables, con distribuciones normales con media 0, varianza 1, siendo algunas de ellas categorizadas para acercarse a un entorno más real. Las variables de exposición/tratamiento y respuesta se definieron de la siguiente manera:

$$p_i = P(T_i = j|X_i) = \frac{e^{\alpha_0 + X_{1,i} + X_{2,i} + X_{3,i} + X_{4,i} + X_{5,i} + X_{6,i} + \varepsilon_i}}{1 + e^{\alpha_0 + X_{1,i} + X_{2,i} + X_{3,i} + X_{4,i} + X_{5,i} + X_{6,i} + \varepsilon_i}}$$

$$\varepsilon_i \sim N(0, 3); \quad j = 0, 1, 2$$

$$Y_i = 1 + X_{1,i} + X_{2,i} + X_{3,i} + X_{4,i} + X_{5,i} + X_{6,i} + T_i + \epsilon_i$$

$$\epsilon_i \sim N(0, 5)$$

La variable de tratamiento se simuló a través de una distribución binomial con parámetros n , t y p_i . La selección del término de error se realizó con fines de agregar ruido a la determinación del estado de los niveles del tratamiento y la variable respuesta.

5.6 Pre procesamiento de los datos

En la simulación de los datos, para cada una de las iteraciones, se calculó los estimadores por mínimos cuadrados ordinarios propios de un modelo de regresión múltiple convencional y se usó la metodología de ponderación media marginal por estratificación (MMWS) (57,58).

Tanto la simulación de los datos, como los posteriores análisis basados en los mismos, se generaron a través del programa estadístico R 3.4.2.

5.7 Consideraciones éticas

El desarrollo de este proyecto no tiene establecido el reclutamiento de pacientes por parte de los investigadores, por lo que según la Resolución 8430 de 1993, que establece las normas científicas, técnicas y administrativas para la investigación en salud, Artículo 11, este estudio pertenece a la categoría de investigación “sin riesgo”. Adicionalmente, no se presentan dilemas éticos inherentes a esta investigación, dada la metodología de la misma. Los resultados aquí obtenidos serán publicados a través de un artículo científico.

Este proyecto fue aprobado por el Comité de Ética de la Facultad de Medicina de la Universidad Nacional de Colombia mediante Acta de Evaluación N.º 016-238-17 del 26 de febrero de 2017 (Anexo 7.1).

5.8 Financiación

El proyecto “Análisis comparativo entre el ajuste por regresión lineal múltiple y Propensity Score ordinal basado en ponderación media Marginal por estratificación (MMWS) en el

control de la confusión en estudios observacionales” fue financiado por recursos propios, usando los recursos bibliográficos de la Universidad Nacional de Colombia.

6. Resultados

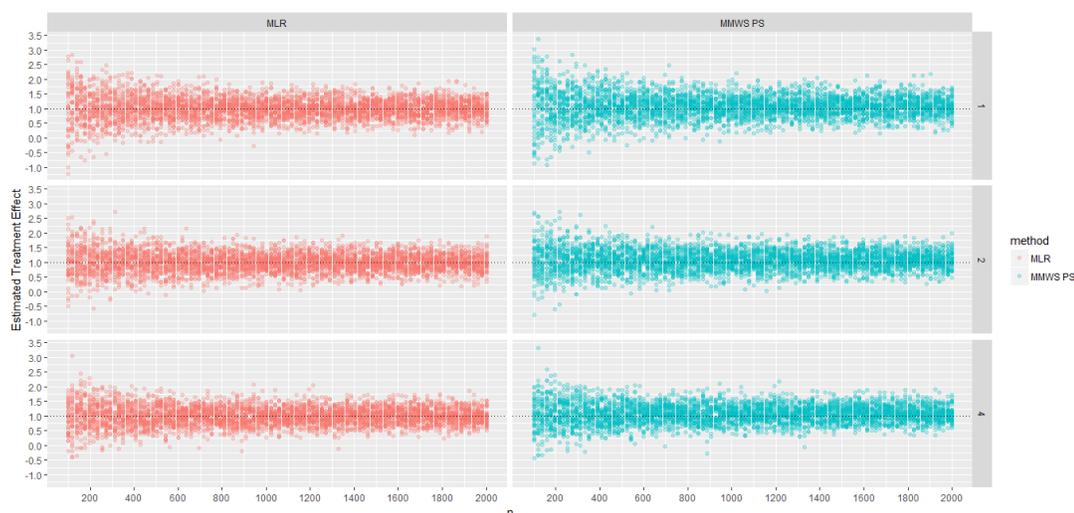
6.1 Resultados de la simulación

6.1.1 Comparación entre regresión lineal y Propensity Score combinado con MMWS

- Comparación en la estimación del primer nivel de tratamiento/exposición (MMWS vs MLR)

La figura 6-1 presenta los resultados de la simulación para el nivel de tratamiento/exposición 1, con tamaños de muestra que van desde 100 individuos hasta 2000, con una razón expuesto: no expuesto en cada grupo de 1, 2 y 4. Se observa que ambos métodos producen estimaciones que son insesgadas para el efecto del tratamiento verdadero, que es igual a 1. Adicionalmente, ambos muestran convergencia hacia el verdadero valor del parámetro cuando el tamaño de muestra aumenta. Es importante resaltar que el rango de las estimaciones para la metodología MMWS son relativamente más grandes en comparación con las estimaciones del modelo de regresión para un tamaño de muestra fijo.

Figura 6-1: Estimaciones del valor del nivel de tratamiento/exposición 1 (MMWS vs MLR)



La comparación del error estándar de la estimación (Figura 6-2), Cuadrado Medio del Error (MSE en inglés) (Figura 6-3), Error Porcentual Absoluto Medio (MAPE en inglés) (Figura 6-4) y la Diferencia Promedio Estandarizada (SMD en inglés) (Figura 6-5), indica que el

Propensity Score ordinal basado en MMWS tiene indicadores de desempeños inferiores a los de una regresión lineal basada en mínimos cuadrados ordinarios para tamaños de muestra inferiores a 500 individuos.

Figura 6-2: Error estándar de la estimación para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)

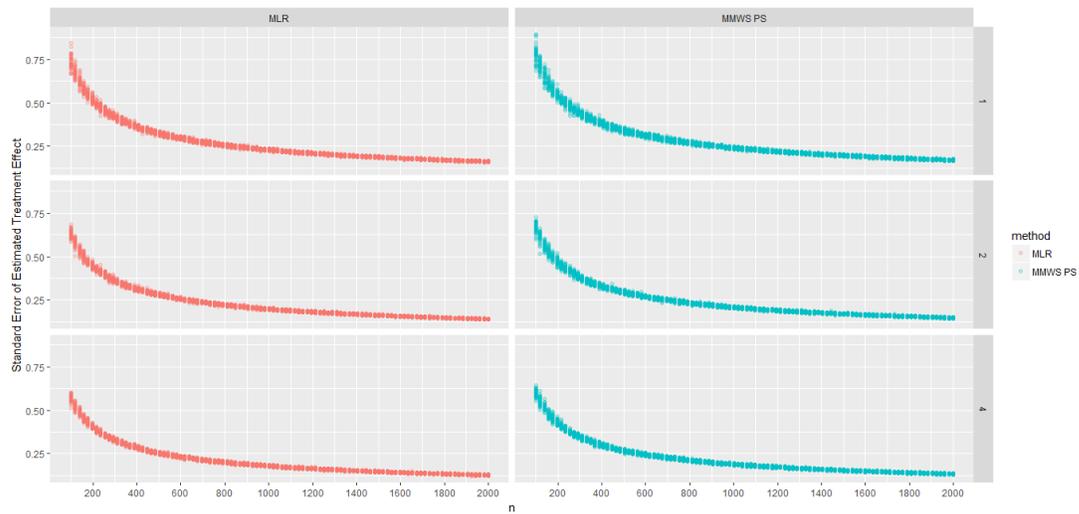


Figura 6-3: Cuadrado Medio del Error para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)

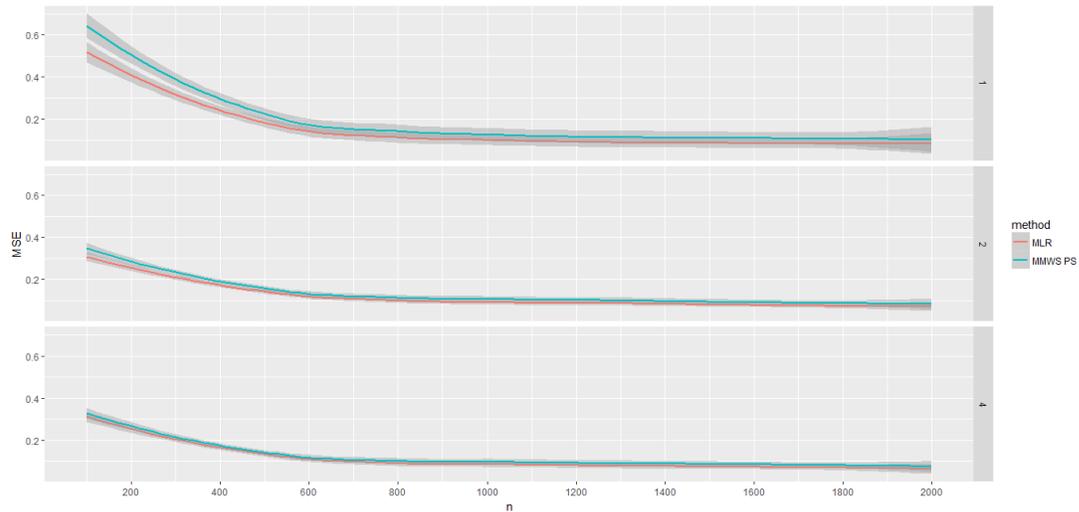


Figura 6-4: Error Porcentual Absoluto Medio para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)

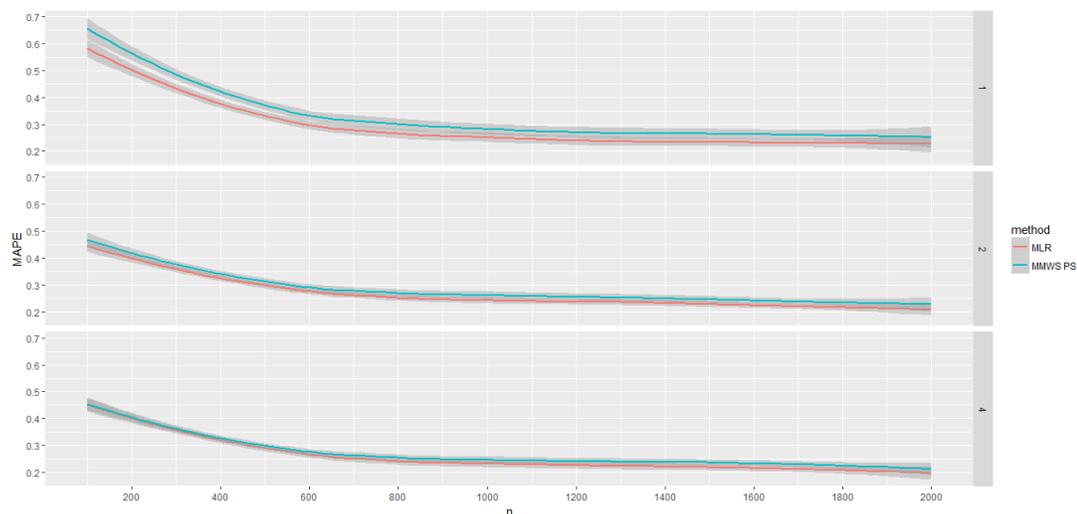
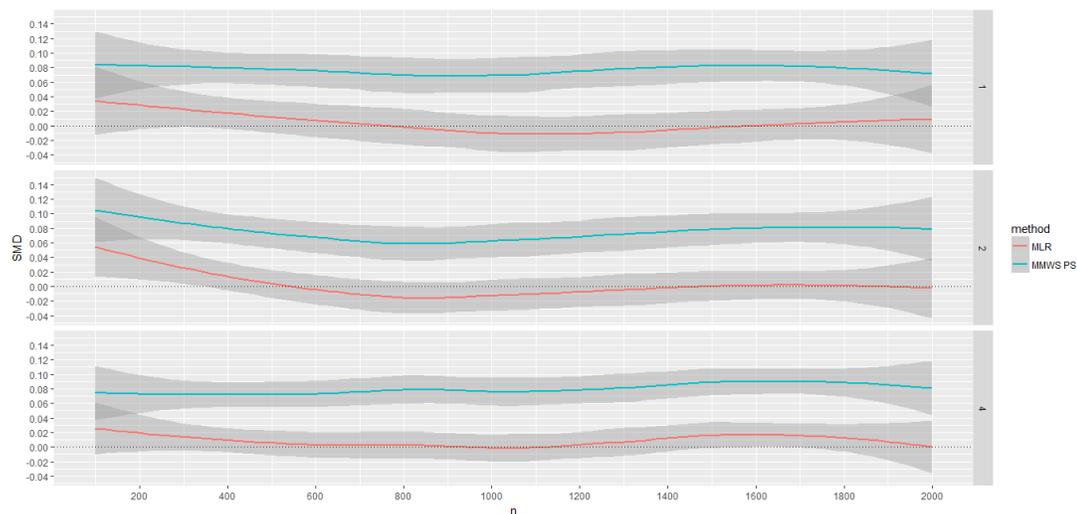
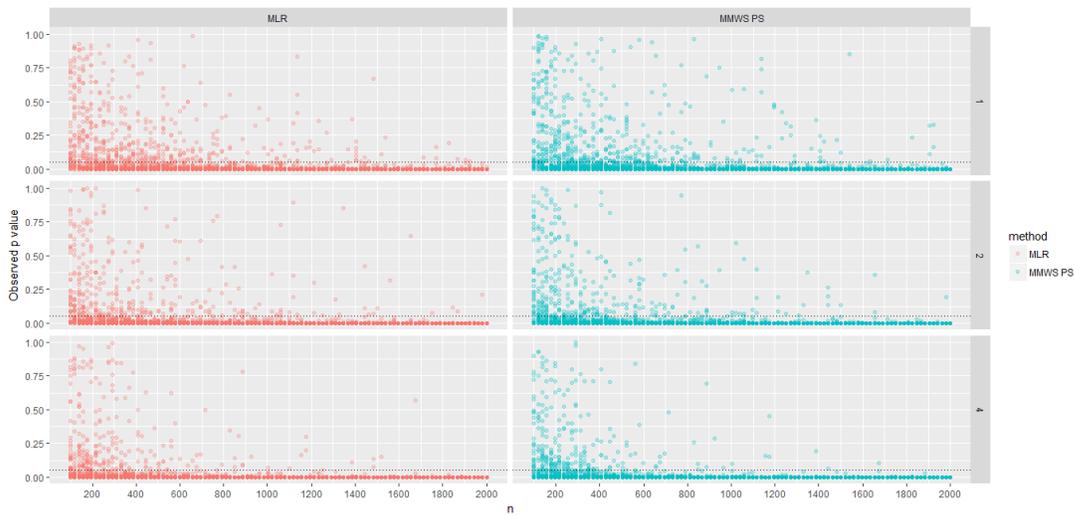
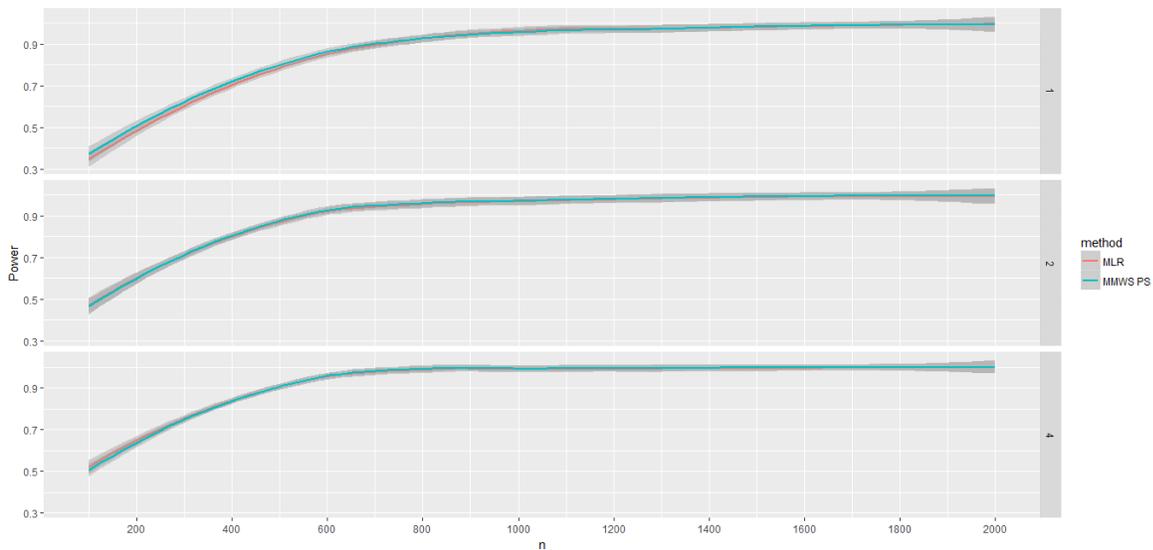


Figura 6-5: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)



Se muestran además de manera complementaria, los valores p observados de los estimadores obtenidos en la simulación y la curva de potencia. Ambos modelos producen resultados similares, sin embargo, la regresión lineal múltiple (MLR) obtiene mejores resultados tal como lo muestran las figuras 6-6 y 6-7.

Figura 6-6: Valor p observado para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)**Figura 6-7:** Curva de potencia para el nivel de tratamiento/exposición 1 (MMWS PS vs MLR)

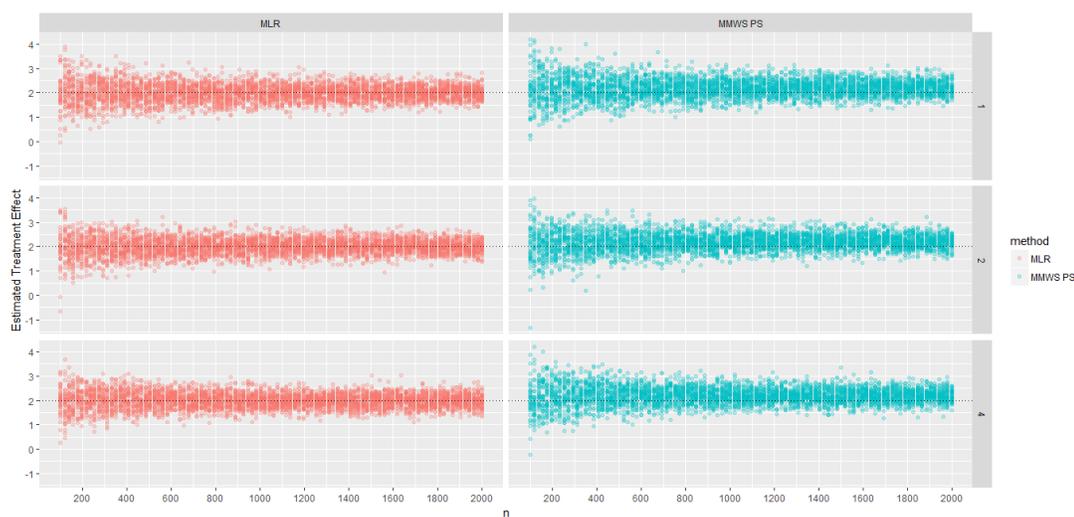
Con base en los anteriores resultados se puede mencionar que las estimaciones basadas en modelos lineales superan a los estimadores MMWS en términos de reducción del sesgo (SMD) de manera importante, para los demás indicadores la diferencia es de alrededor de 5 puntos porcentuales (MSE, MAPE), sin embargo, esta diferencia se reduce de manera notable para tamaños de muestra superiores a 500 individuos. Cabe resaltar que en términos de las estimaciones del efecto y sus correspondientes errores estándar las

diferencias entre ambas metodologías son mínimas, situación análoga se presenta con los valores p observados y las curvas de potencia de los modelos comparados.

▪ **Comparación en la estimación del segundo nivel de tratamiento/exposición (MMWS vs MLR)**

Con respecto al segundo nivel de tratamiento/exposición, se muestran los indicadores de desempeño, donde se realizó una comparación entre los estimadores dados por una regresión lineal múltiple convencional en contra del Propensity Score basado en MMWS, resalta la convergencia al verdadero valor del parámetro (dos en este caso) para ambos modelos, tal como se expone en la figura 6-8.

Figura 6-8: Estimaciones del valor del nivel de tratamiento/exposición 2 (MMWS PS vs MLR)



Los indicadores de desempeño para los modelos a comparar: Error estándar de la estimación (Figura 6-9), Cuadrado Medio del Error (MSE) (Figura 6-10), Error Porcentual Absoluto Medio (MAPE) (Figura 6-11) y la Diferencia Promedio Estandarizada (SMD) (Figura 6-12), muestran un desempeño inferior del Propensity Score ordinal basado en MMWS en comparación con una regresión lineal múltiple basada en mínimos cuadrados ordinarios.

Figura 6-9: Error estándar de la estimación para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)

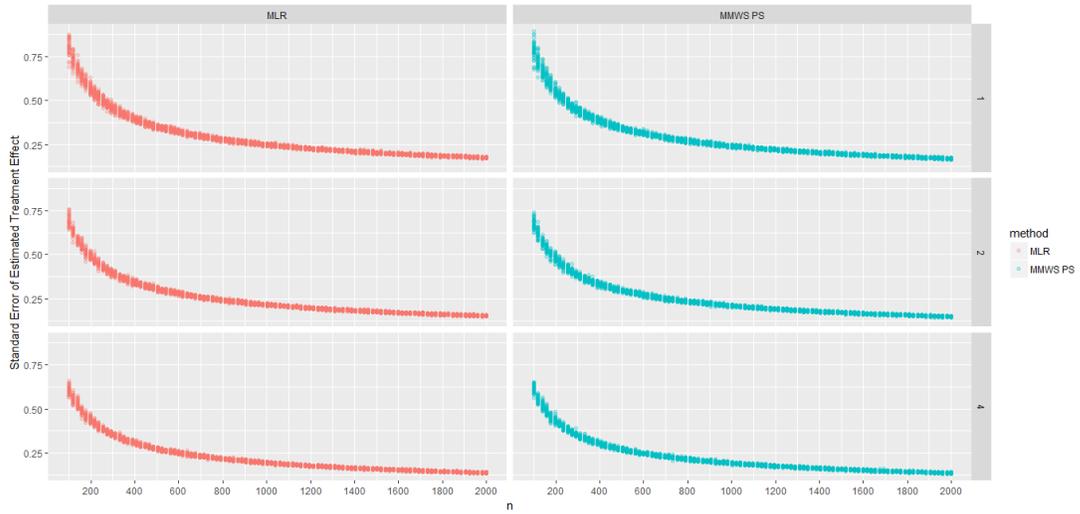


Figura 6-10: Cuadrado Medio del Error para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)

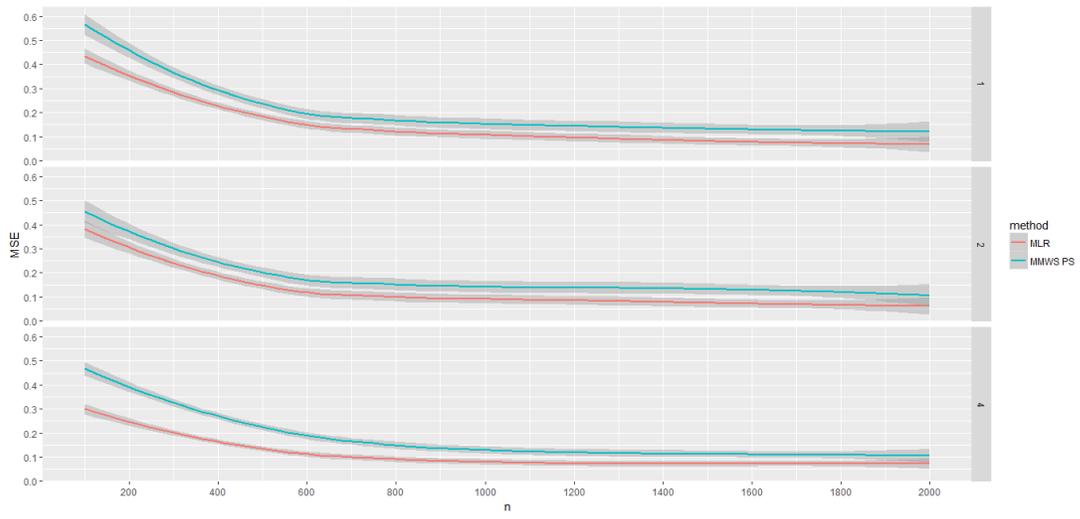


Figura 6-11: Error Porcentual Absoluto Medio para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)

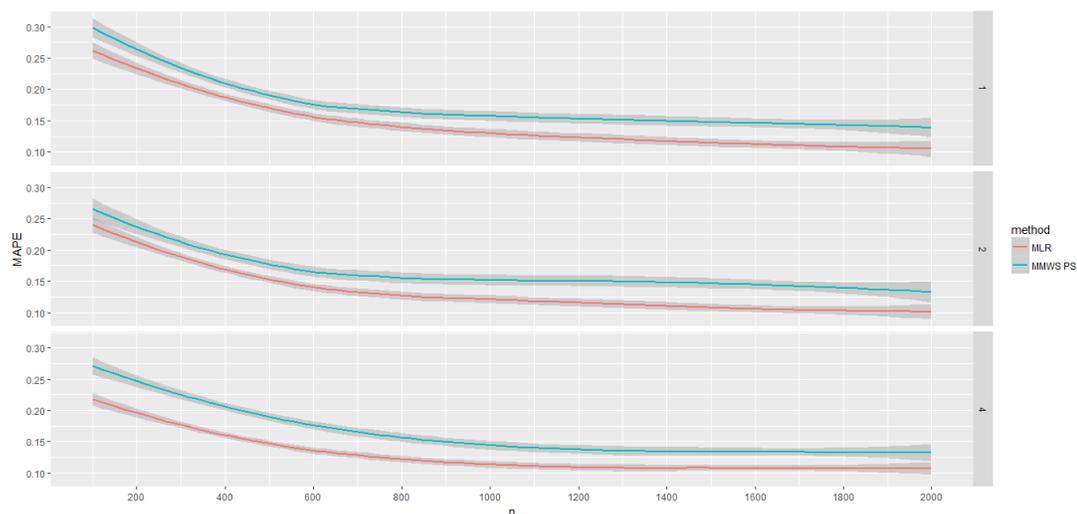
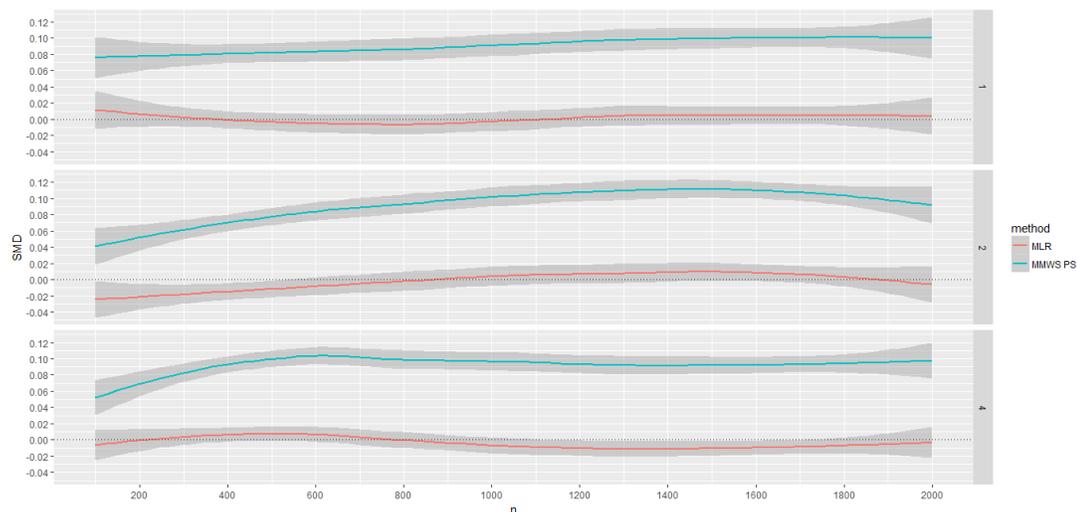


Figura 6-12: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)



Los resultados para el segundo nivel de tratamiento/exposición son similares a los mencionados en el apartado anterior; convergencia hacia el verdadero valor del parámetro (en este caso un valor igual a dos) a medida que aumenta el tamaño de muestra y la razón expuesto: no expuesto aumenta. Se observa una diferencia importante entre la precisión de los estimadores a favor de la regresión lineal múltiple en contra del Propensity Score basado en MMWS.

Adicionalmente, las figuras 6-13 y 6-14 exhiben los valores p observados de los estimadores obtenidos en la simulación y la correspondiente curva de potencia para este segundo nivel de tratamiento/exposición.

Figura 6-13: Valor p observado para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)

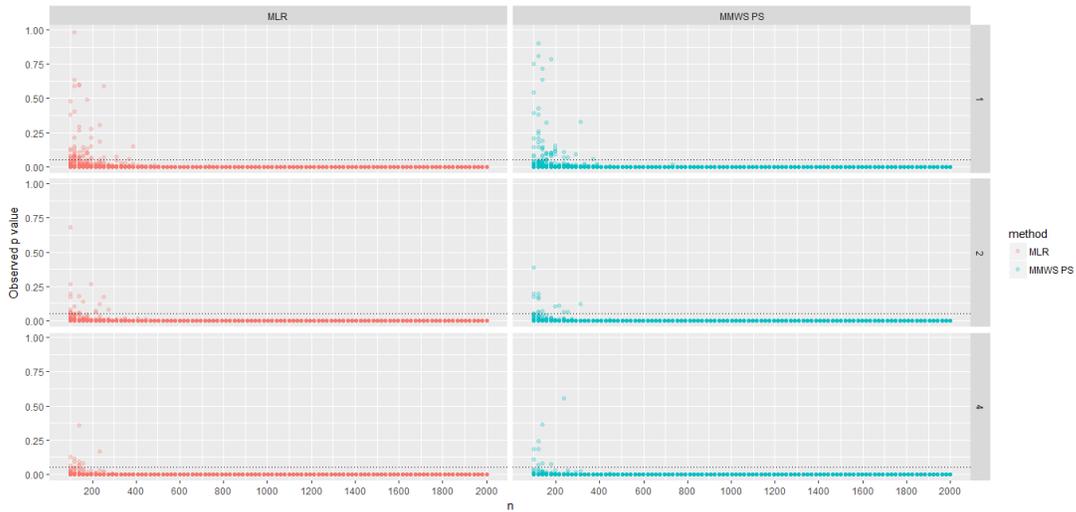
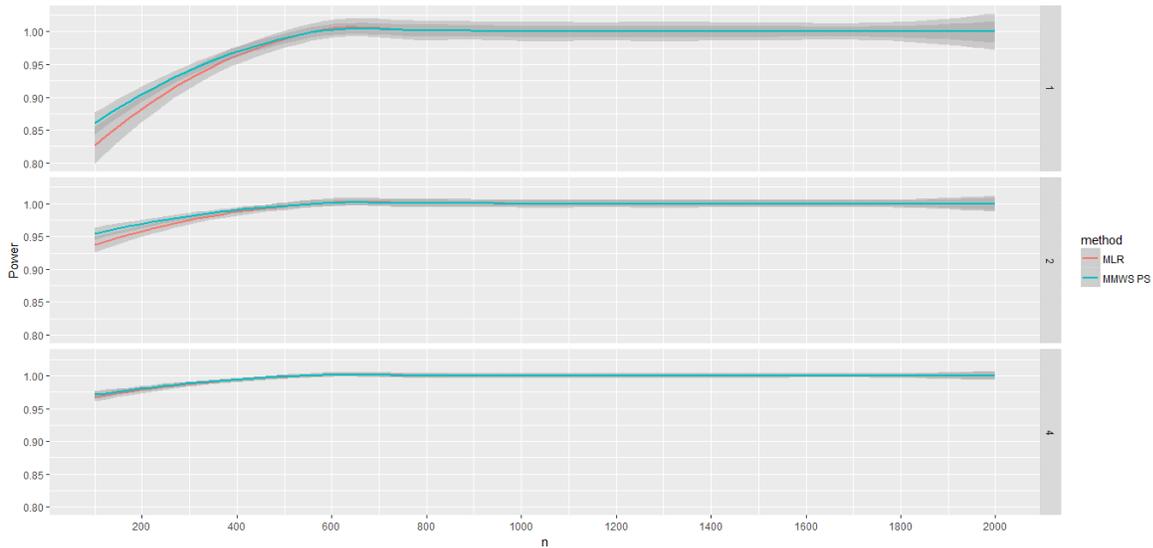


Figura 6-14: Curva de potencia para el nivel de tratamiento/exposición 2 (MMWS PS vs MLR)



Las conclusiones obtenidas para el primer nivel de tratamiento/exposición son extrapolables a este segundo nivel de tratamiento/exposición: convergencia de los estimadores al verdadero valor del parámetro de manera más rápida cuando se aumenta la razón no expuesto: expuesto; diferencias cada vez más reducidas entre las

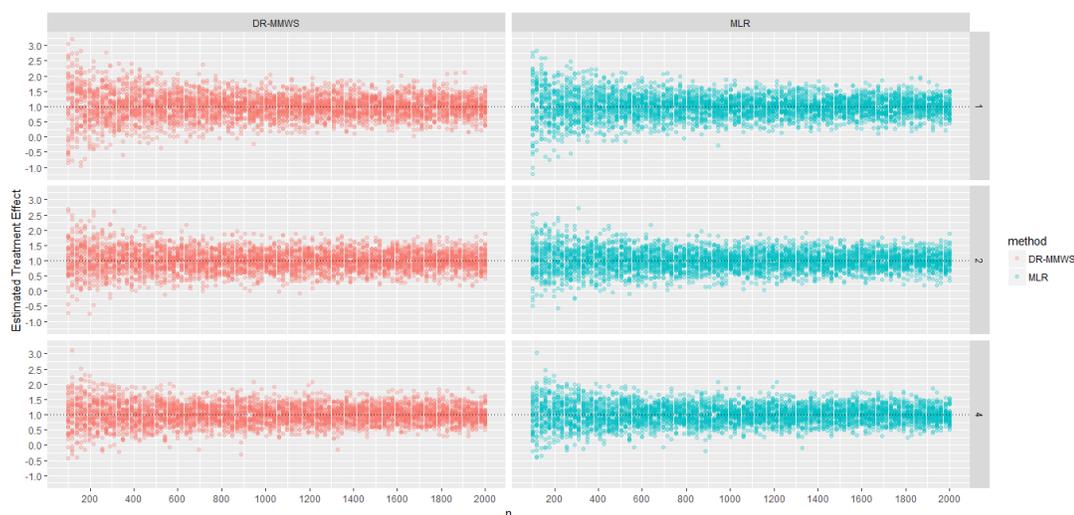
metodologías comparadas ante un aumento del tamaño de muestra. Por último, los estimadores basados en modelos de regresión lineal son superiores a los estimadores basados en MMWS para un tamaño de muestra y números de variables fijos.

6.1.2 Propiedades de los estimadores doblemente robustos (DR-MMWS)

▪ Comparación en la estimación del primer nivel de tratamiento/exposición (DR-MMWS vs MLR)

Con respecto a los estimadores doblemente robustos (DR-MMWS), los resultados se muestran a continuación. Cabe resaltar que, en cuanto a la estimación del tamaño del efecto, se observa un comportamiento similar entre los estimadores doblemente robustos y una regresión lineal múltiple (MLR), convergencia de los estimadores al verdadero valor del parámetro con mayores tamaños de muestra y con resultados insesgados (Figura 6-15).

Figura 6-15: Estimaciones del valor del nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)



La comparación del error estándar de la estimación (Figura 6-16), Cuadrado Medio del Error (MSE) (Figura 6-17), Error Porcentual Absoluto Medio (MAPE) (Figura 6-18) y la Diferencia Promedio Estandarizada (SMD) (Figura 6-19), sugieren que el Propensity Score ordinal basado en DR-MMWS tiene indicadores de desempeños similares a los obtenidos por una regresión lineal múltiple para un número de variables establecido y con un tamaño

de muestra variable. Para tamaños de muestra superiores a 500 los resultados son prácticamente iguales.

Figura 6-16: Error estándar de la estimación para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)

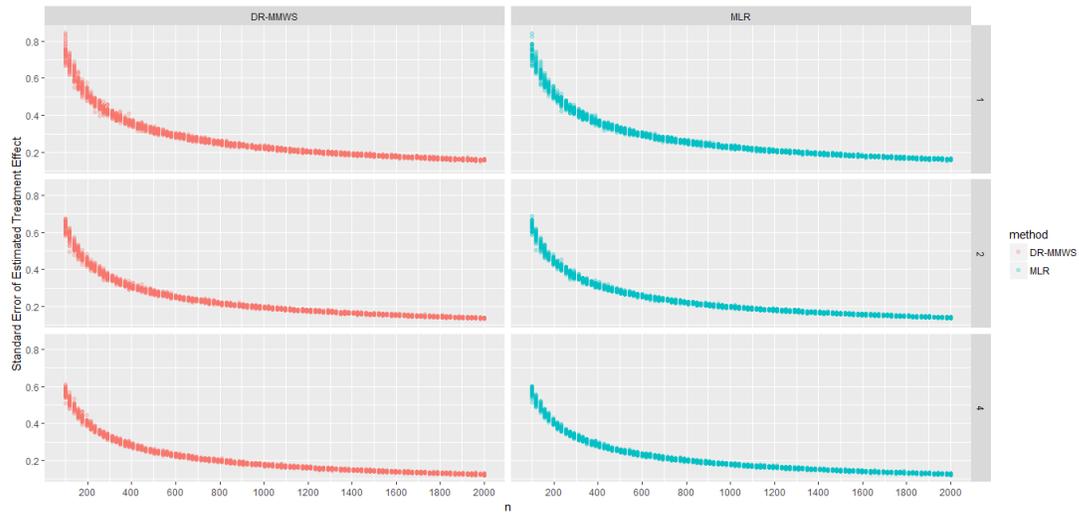


Figura 6-17: Cuadrado Medio del Error del nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)

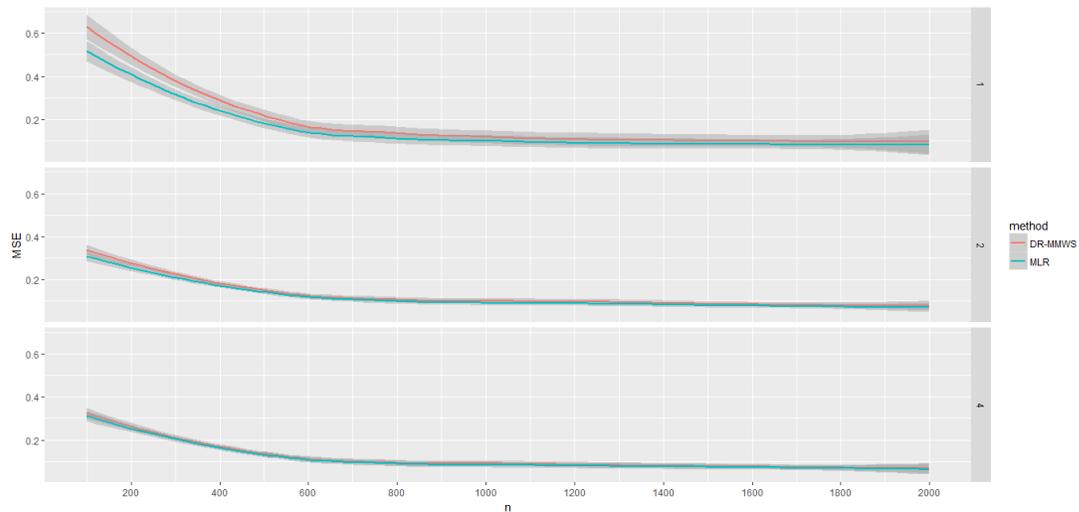


Figura 6-18: Error Porcentual Absoluto Medio del nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)

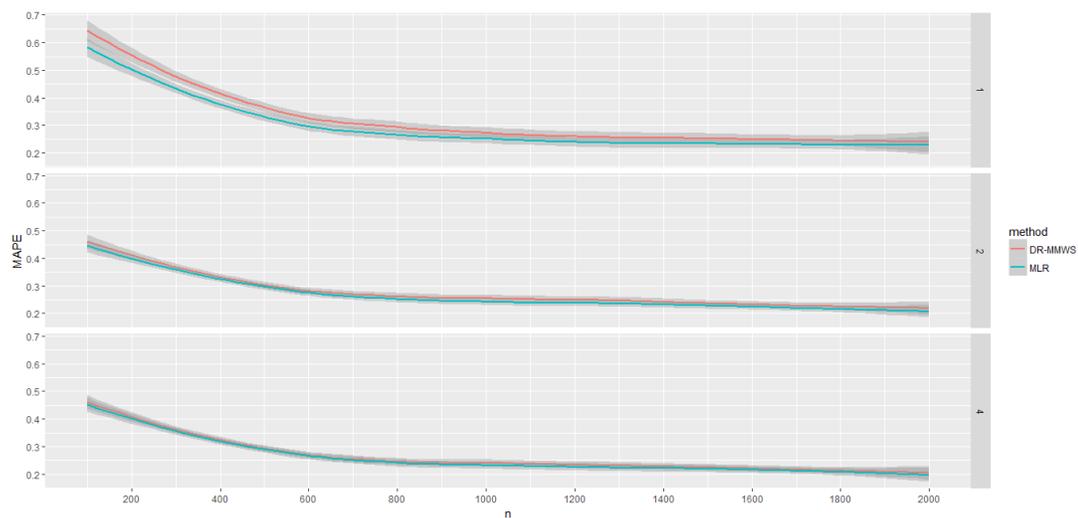


Figura 6-19: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)

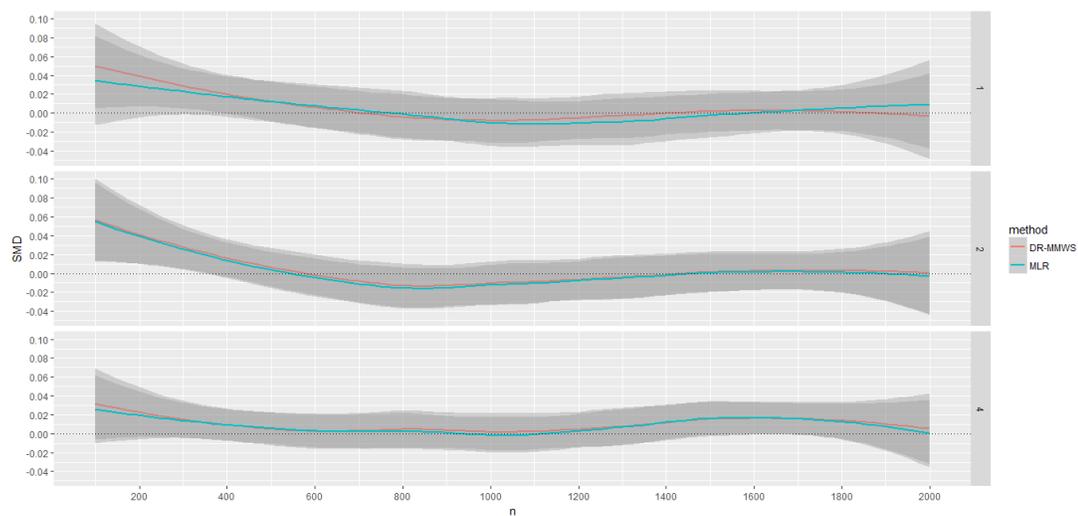


Figura 6-20: Valor p observado para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)

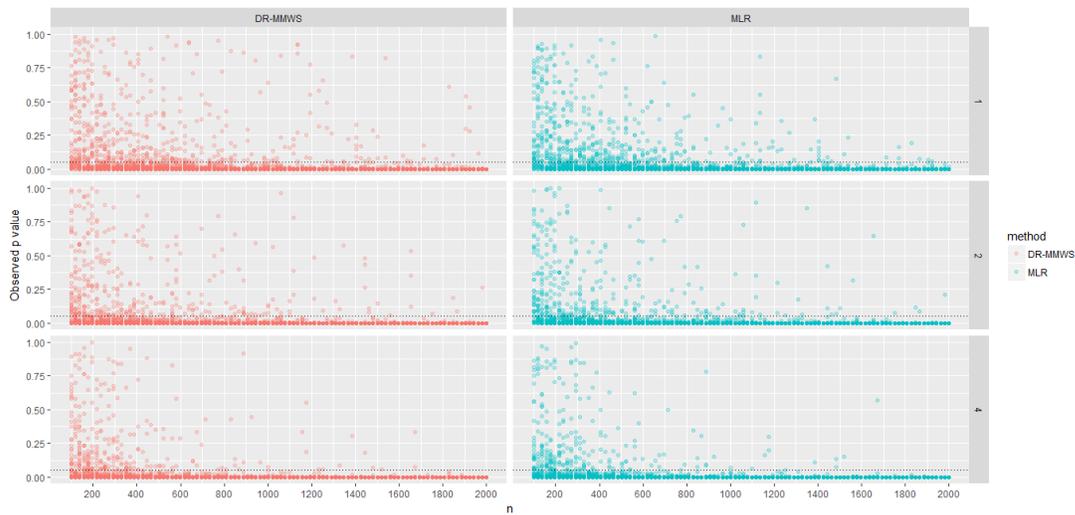
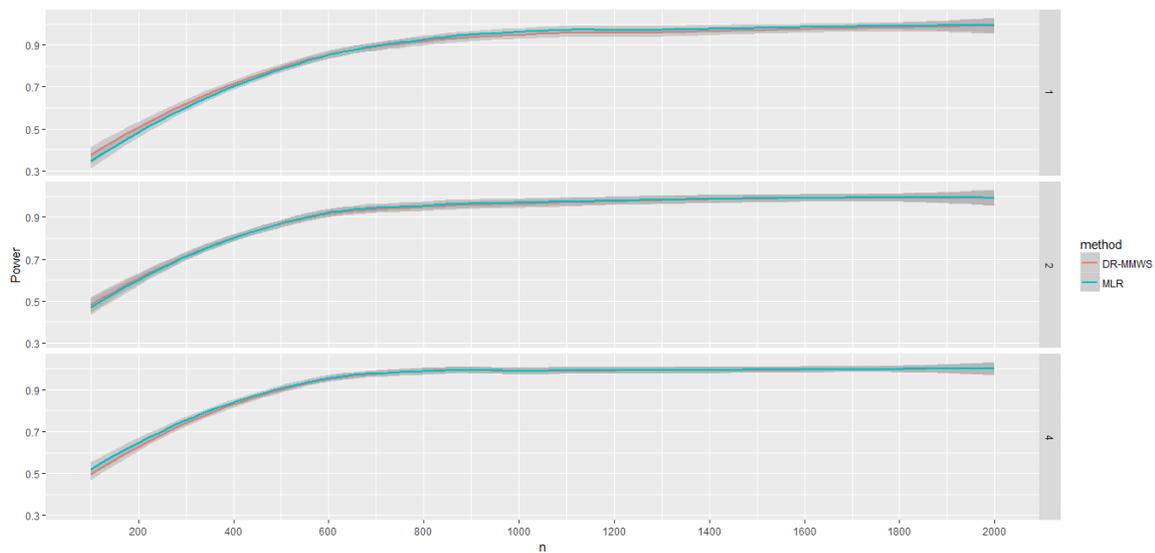


Figura 6-21: Curva de potencia para el nivel de tratamiento/exposición 1 (DR-MMWS vs MLR)

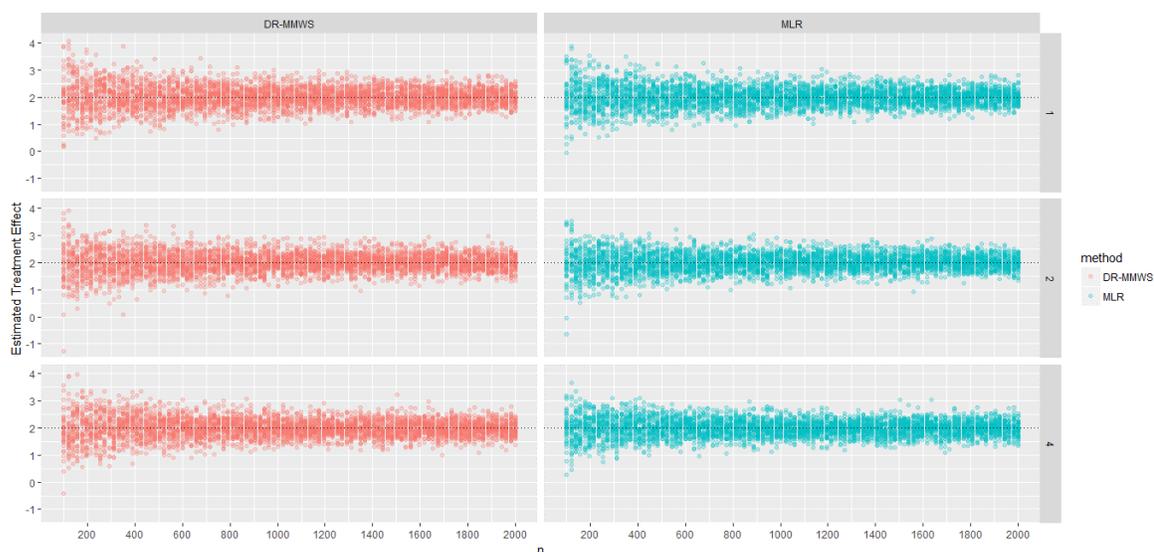


Se muestran de manera complementaria, los valores p observados de los estimadores obtenidos en la simulación (Figura 6-20) y la curva de potencia correspondiente al primer nivel de tratamiento/exposición (Figura 6-21). Como se puede observar, ambos modelos producen resultados similares, tanto en los valores p observados como en la curva de potencia.

▪ **Comparación en la estimación del segundo nivel de tratamiento/exposición (DR-MMWS vs MLR)**

Como se puede observar en la siguiente figura (Figura 6-22) para este nivel de tratamiento/exposición las estimaciones y errores estándar de ambos métodos son similares.

Figura 6-22: Estimaciones del valor del nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)



En cuanto al error estándar de la estimación (Figura 6-23), Cuadrado Medio del Error (MSE) (Figura 6-24), Error Porcentual Absoluto Medio (MAPE) (Figura 6-25) y la Diferencia Promedio Estandarizada (SMD) (Figura 6-26), tanto la regresión por mínimos cuadrados ordinarios como el estimador doblemente robusto DR-MMWS tienen indicadores de desempeños semejantes.

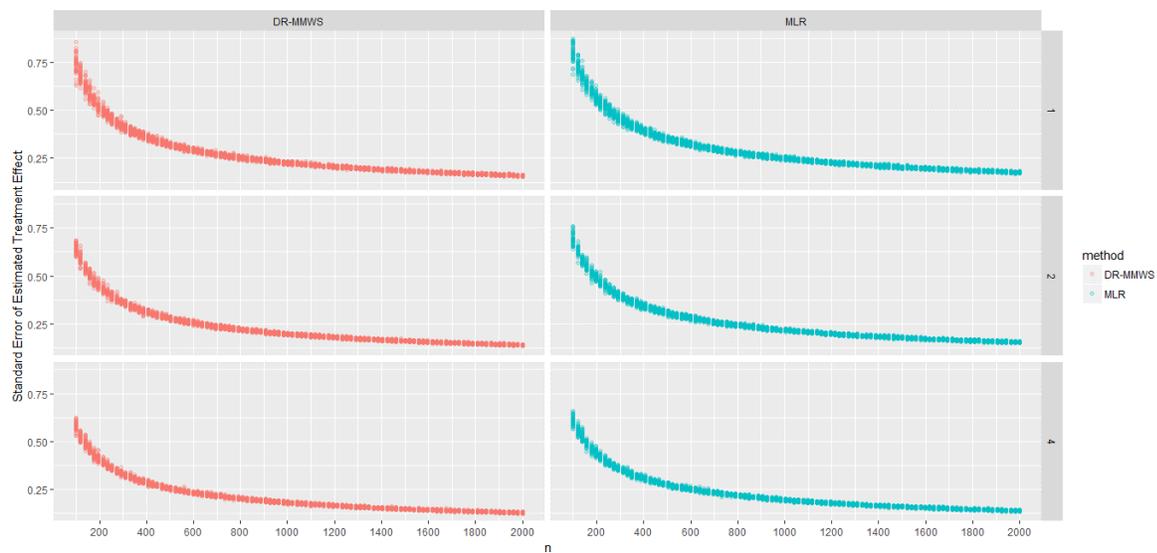
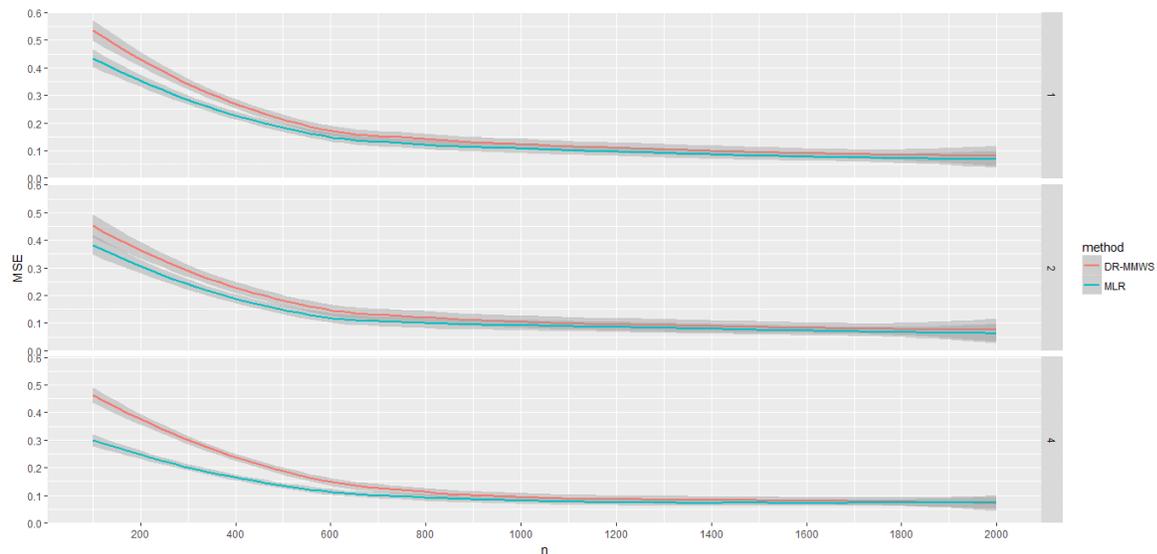
Figura 6-23: Error estándar de la estimación para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)**Figura 6-24:** Cuadrado Medio del Error del nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)

Figura 6-25: Error Porcentual Absoluto Medio del nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)

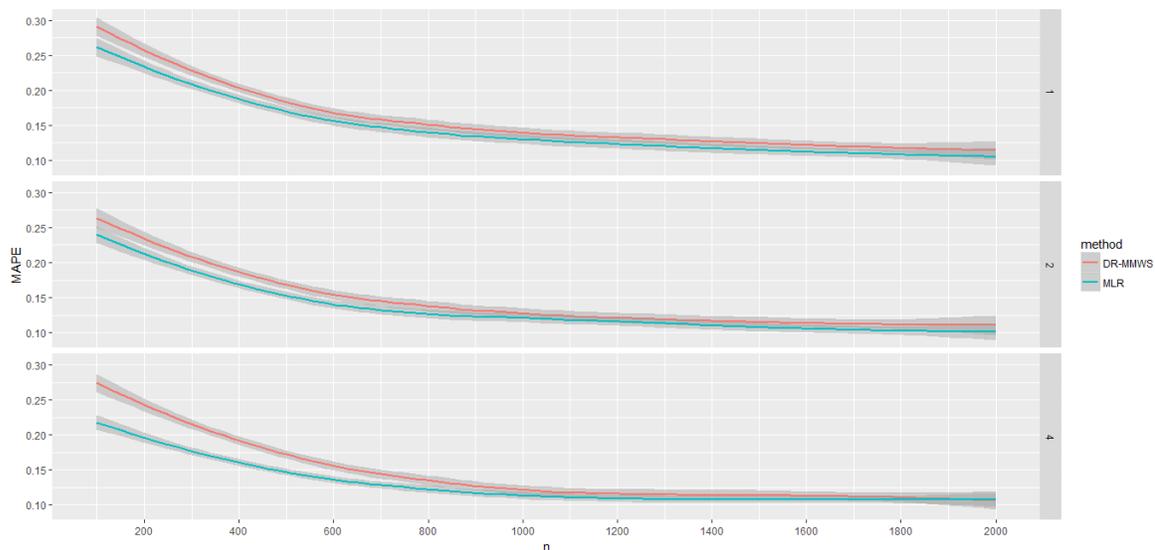
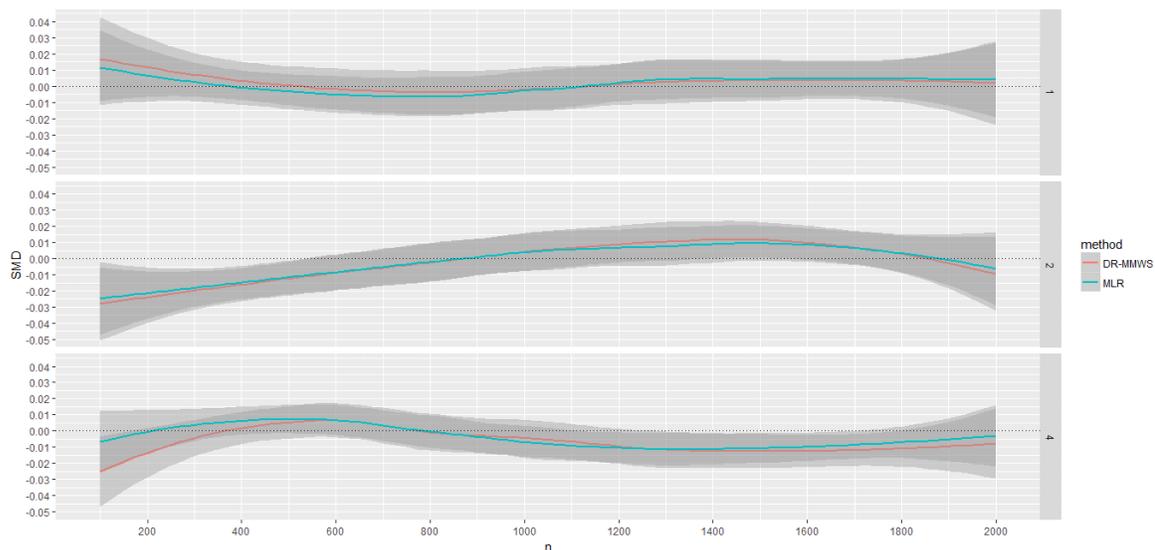
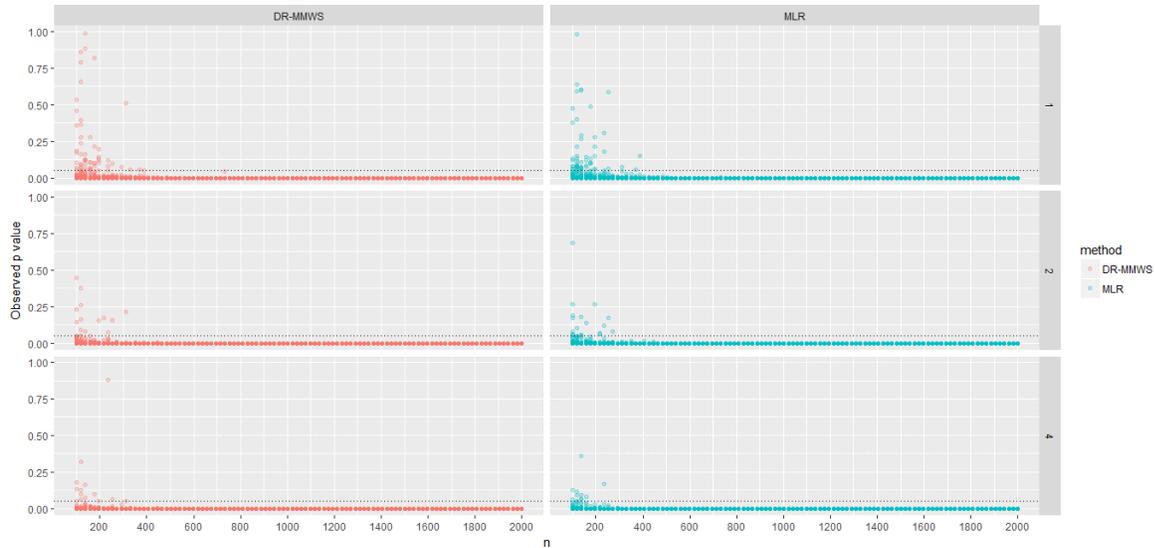
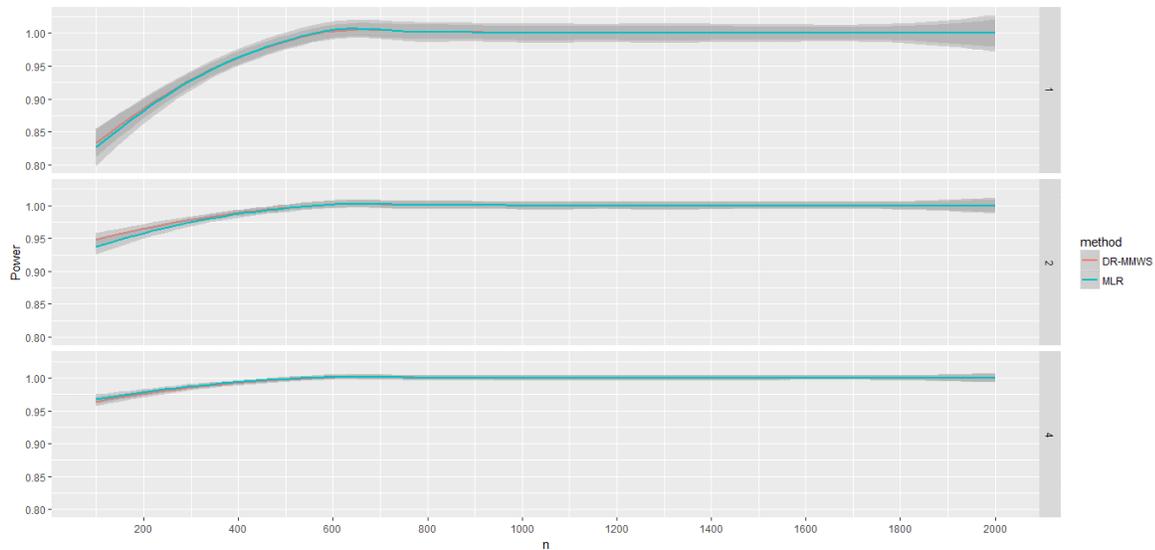


Figura 6-26: Diferencia promedio estandarizada para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)



Se muestran además de manera complementaria, los valores p observados de los estimadores obtenidos en la simulación (Figura 6-27) y la curva de potencia (Figura 6-28). Ambos modelos producen resultados similares, aunque, al igual que con los gráficos estimados y MSE anteriores, fenómeno similar surge al comparar la curva de potencia entre los métodos en cuestión.

Figura 6-27: Valor p observado para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)**Figura 6-28:** Curva de potencia para el nivel de tratamiento/exposición 2 (DR-MMWS vs MLR)

Con base en los anteriores resultados, se puede resaltar que las estimaciones basadas en un Propensity Score ordinal combinado con una ponderación media Marginal por estratificación (DR-MMWS) son levemente inferiores en términos de reducción del sesgo a los obtenidos a través de una regresión lineal múltiple, presenta un desempeño análogo a los estimadores de regresión basados en mínimos cuadrados ordinarios, en especial ante un aumento en el tamaño de muestra o la relación expuesto: no expuesto.

6.1.3 Efecto de un mayor número de variables

Con el fin de evaluar el efecto que tiene la inclusión de una nueva variable en los dos modelos comparados, se realiza un análisis de simulación: con un tamaño de muestra de 500 individuos, se aplican cada uno de los métodos anteriormente descritos. La figura 6-29 presenta los resultados de la simulación para el nivel de tratamiento/exposición 1, el número de variables incluidas van desde 6 hasta 11, con una razón no expuesto: expuesto en cada grupo de 1, 2 y 4. Los métodos en cuestión producen estimaciones insesgadas para un efecto del tratamiento verdadero, que es igual a 1, aunque no se evidencia alguna ganancia en términos de reducción de la variabilidad del efecto estimado con un mayor número de variables (Figura 6-29). Con respecto a los errores estándar se observa un aumento en su magnitud con la inclusión de más variables, siendo especialmente visibles en los estimadores MMWS (Figura 6-30).

Figura 6-29: Comparación de las estimaciones del valor del nivel de tratamiento/ 1 (DR-MMWS, MLR y MMWS PS)

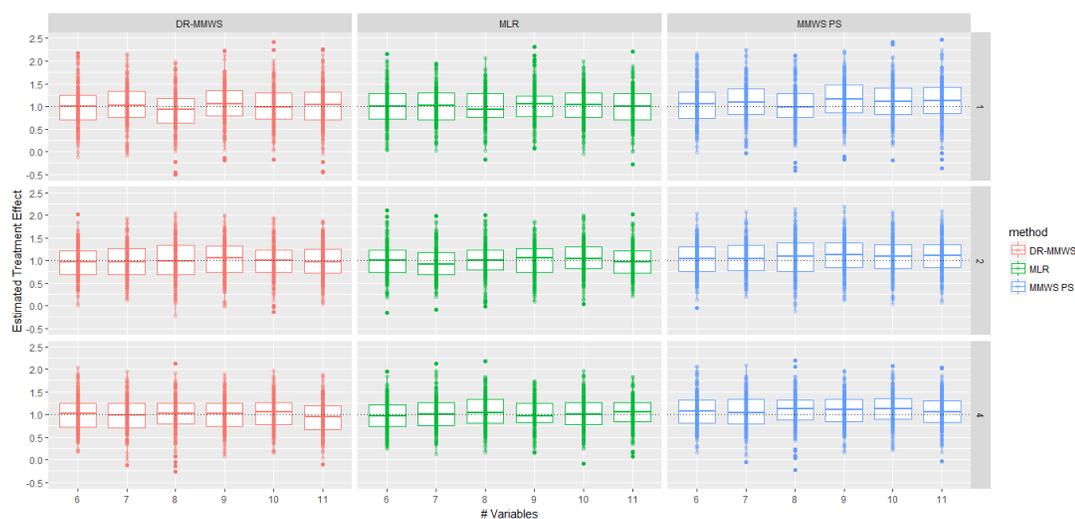
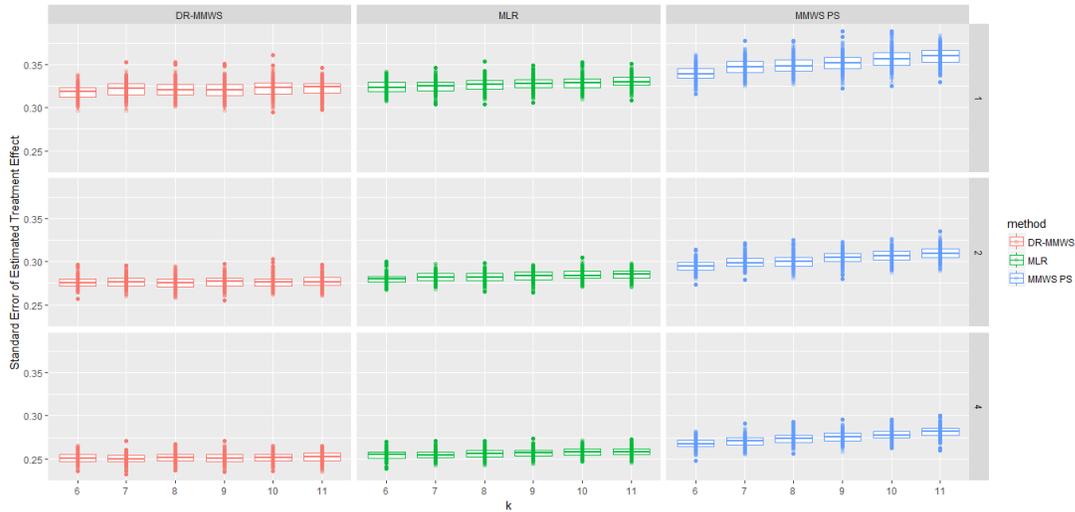


Figura 6-30: Comparación de los Errores estándar de la estimación para el nivel de tratamiento/exposición 1 (DR-MMWS, MLR y MMWS PS)



A continuación, se muestra los indicadores de desempeño para los tres métodos comparados. Se obtienen para cada uno de los modelos el efecto estimado para el tratamiento/exposición 1, Error Estándar Promedio, Cuadrado Medio del Error (MSE) y Diferencia Promedio Estandarizada (SMD) con el número de variables mencionados anteriormente (Tabla 6-1). Se observa que los indicadores de desempeño no presentan cambios significativos con la inclusión de más variables. La relación en el número de no expuestos por expuestos, por el contrario, disminuye el Error Estándar de la estimación, el Cuadrado Medio del Error (MSE) y la Diferencia Promedio Estandarizada (SMD). Un mayor número de sujetos no expuestos por sujetos expuestos mejora la precisión de los estimadores independiente del número de variables utilizada en los modelos comparados.

Tabla 6-1: Resultados de la simulación de Monte Carlo para los métodos comparados con diversos números de variables predictoras para el tratamiento/exposición 1

Modelo	Ratio	k	Efecto estimado	Error Estándar	MSE	SDM
DR-MMWS	1:1	6	1,000	0,318	0,198	0,00%
		7	1,026	0,321	0,184	2,59%
		8	0,921	0,321	0,205	-7,87%
		9	1,051	0,321	0,187	5,14%
		10	1,010	0,323	0,189	1,03%
		11	1,013	0,322	0,221	1,32%
	1:2	6	0,959	0,276	0,131	-4,14%
		7	0,981	0,277	0,141	-1,89%
		8	0,990	0,276	0,192	-1,02%
		9	1,023	0,277	0,162	2,34%
		10	0,994	0,277	0,155	-0,64%
		11	0,991	0,277	0,151	-0,86%
	1:4	6	1,010	0,251	0,143	1,03%
		7	0,983	0,250	0,137	-1,67%
		8	1,008	0,251	0,142	0,78%
		9	0,996	0,251	0,124	-0,42%
		10	1,030	0,252	0,129	3,04%
		11	0,952	0,252	0,131	-4,80%
MLR	1:1	6	1,007	0,324	0,176	0,71%
		7	1,012	0,325	0,162	1,24%
		8	0,980	0,327	0,150	-1,99%
		9	1,041	0,328	0,149	4,09%
		10	1,020	0,329	0,154	2,03%
		11	1,002	0,330	0,163	0,22%
	1:2	6	0,990	0,280	0,141	-1,03%
		7	0,948	0,282	0,133	-5,19%
		8	1,005	0,282	0,133	0,54%
		9	0,995	0,284	0,137	-0,51%
		10	1,044	0,285	0,135	4,36%
		11	0,986	0,285	0,116	-1,43%
	1:4	6	0,975	0,255	0,116	-2,47%
		7	1,017	0,255	0,120	1,68%
		8	1,067	0,256	0,140	6,70%
		9	1,015	0,257	0,106	1,53%
		10	1,020	0,258	0,129	2,04%
		11	1,035	0,258	0,104	3,55%

MMWS	1:1	6	1,065	0,339	0,210	6,51%
		7	1,103	0,347	0,195	10,27%
		8	1,010	0,349	0,203	1,03%
		9	1,145	0,352	0,209	14,54%
		10	1,111	0,357	0,200	11,06%
		11	1,115	0,360	0,228	11,54%
	1:2	6	1,025	0,295	0,133	2,47%
		7	1,056	0,299	0,140	5,65%
		8	1,072	0,301	0,202	7,20%
		9	1,116	0,305	0,167	11,59%
		10	1,090	0,307	0,163	9,01%
		11	1,094	0,310	0,164	9,42%
	1:4	6	1,080	0,268	0,148	8,00%
		7	1,058	0,271	0,139	5,80%
		8	1,096	0,274	0,146	9,56%
		9	1,087	0,276	0,129	8,66%
		10	1,124	0,278	0,144	12,39%
		11	1,058	0,282	0,132	5,83%

Para el nivel de tratamiento/exposición 2, en la figura 6-31 se observan los resultados de la simulación. Se observa que ambos estimadores producen estimaciones insesgadas para un efecto del tratamiento verdadero, que es igual a 2. Se presentan los mismos resultados que en el nivel anterior, no se observa una reducción en la variabilidad del efecto estimado con un mayor número de variables (Figura 6-31). Por otro lado, los errores estándar aumentan en el error estándar con la inclusión de más variables, los estimadores doblemente robustos DR-MMWS presentaron el menor cambio de los tres métodos a comparar (Figura 6-32).

Figura 6-31: Comparación de las estimaciones del valor del nivel de tratamiento/ 2 (DR-MMWS, MLR y MMWS PS)

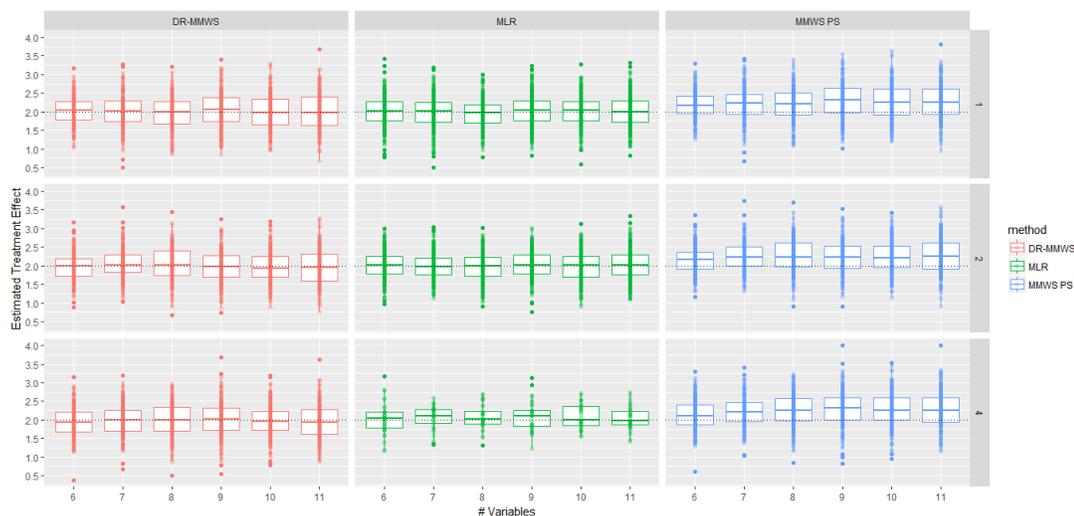
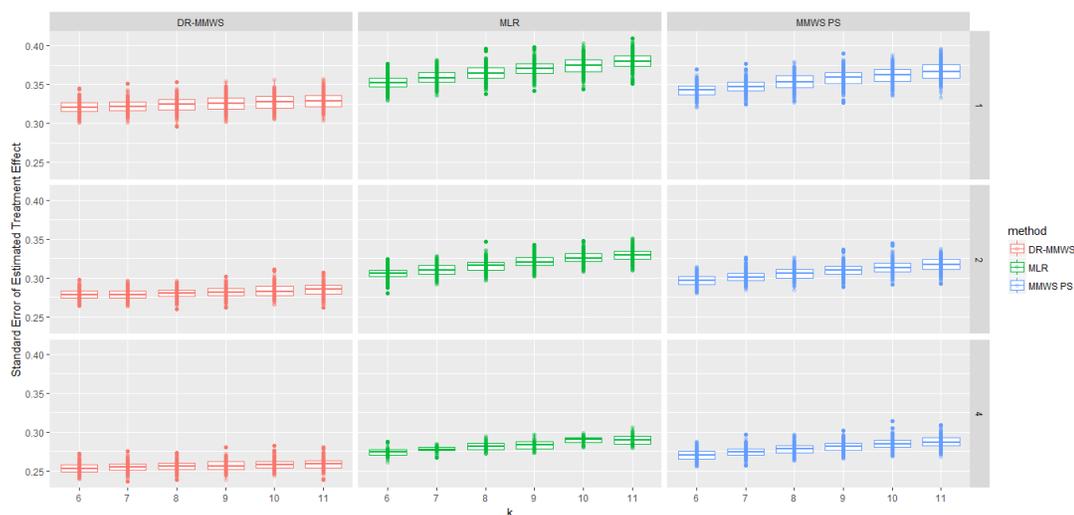


Figura 6-32: Comparación de los Errores estándar de la estimación para el nivel de tratamiento/exposición 1 (DR-MMWS, MLR y MMWS PS)



Análogo a lo presentado anteriormente, se obtienen para cada uno de los modelos el efecto estimado para el tratamiento/exposición 2, Error Estándar Promedio, Cuadrado Medio del Error (MSE) y Diferencia Promedio Estandarizada (SMD) con el número de variables que se definió el inicio de esta sección (Tabla 6-2). Según los datos descritos en la tabla 6-2, no se manifiesta variaciones importantes en los indicadores de desempeño con la inclusión de más variables. De nuevo, una relación de no expuestos: expuestos mayor, conduce a una mejoría en los indicadores de desempeño.

Tabla 6-2: Resultados de la simulación de Monte Carlo para los métodos comparados con diversos números de variables predictoras para el tratamiento/exposición 2

Modelo	Ratio	k	Efecto estimado	Error Estándar	MSE	SDM
DR-MMWS	1:1	6	2,024	0,321	0,152	1,19%
		7	2,015	0,322	0,200	0,75%
		8	1,968	0,324	0,203	-1,61%
		9	2,049	0,327	0,251	2,43%
		10	2,001	0,328	0,236	0,05%
		11	2,003	0,330	0,252	0,14%
	1:2	6	1,973	0,278	0,148	-1,36%
		7	2,056	0,279	0,169	2,79%
		8	2,042	0,280	0,210	2,10%
		9	1,970	0,281	0,199	-1,52%
		10	1,961	0,284	0,194	-1,97%
		11	1,982	0,284	0,233	-0,90%
	1:4	6	1,941	0,253	0,177	-2,97%
		7	1,991	0,255	0,173	-0,43%
		8	1,998	0,256	0,208	-0,08%
		9	2,016	0,257	0,228	0,81%
		10	1,975	0,258	0,196	-1,23%
		11	1,947	0,259	0,234	-2,64%
MLR	1:1	6	2,009	0,353	0,163	0,45%
		7	1,987	0,359	0,187	-0,65%
		8	1,966	0,365	0,165	-1,68%
		9	2,030	0,371	0,189	1,50%
		10	2,021	0,375	0,163	1,04%
		11	1,997	0,380	0,193	-0,16%
	1:2	6	2,012	0,306	0,130	0,59%
		7	1,994	0,310	0,119	-0,28%
		8	1,975	0,316	0,147	-1,25%
		9	2,016	0,321	0,129	0,78%
		10	1,987	0,326	0,152	-0,64%
		11	2,030	0,329	0,166	1,50%
	1:4	6	2,028	0,274	0,185	1,41%
		7	2,055	0,277	0,102	2,73%
		8	2,083	0,282	0,116	4,14%

		9	2,102	0,283	0,170	5,12%
		10	2,099	0,289	0,117	4,93%
		11	2,043	0,290	0,096	2,15%
MMWS	1:1	6	2,187	0,343	0,177	9,35%
		7	2,204	0,348	0,243	10,21%
		8	2,184	0,354	0,235	9,22%
		9	2,283	0,358	0,319	14,13%
		10	2,258	0,362	0,307	12,91%
		11	2,273	0,367	0,317	13,65%
	1:2	6	2,153	0,297	0,169	7,65%
		7	2,260	0,301	0,229	13,02%
		8	2,269	0,306	0,275	13,45%
		9	2,229	0,309	0,245	11,44%
		10	2,241	0,314	0,248	12,07%
		11	2,271	0,317	0,304	13,55%
	1:4	6	2,125	0,270	0,187	6,26%
		7	2,219	0,274	0,216	10,94%
		8	2,245	0,278	0,260	12,25%
		9	2,288	0,281	0,309	14,38%
		10	2,276	0,285	0,292	13,78%
		11	2,274	0,287	0,312	13,69%

7. Discusión

En este estudio, se utilizaron simulaciones de Monte Carlo para comparar el rendimiento del estimador MMWS con un ajuste de regresión lineal múltiple convencional, metodologías usadas para estimar los efectos del tratamiento en estudios observacionales. El estudio busco determinar en qué condiciones de tamaño de muestra, relación expuesto-no expuesto y número de variables los estimadores basado en MMWS en una exposición/tratamiento de tipo ordinal otorgan una estimación del tamaño del efecto más preciso e insesgado. Los predictores continuos utilizados provenían de distribuciones normales y para las variables categóricas se colapsaron dos distribuciones normales de dos y tres categorías respectivamente, estos fueron los únicos elementos de ruido no blanco en la determinación del estado del tratamiento. Adicionalmente el tratamiento/exposición está basado en un modelo lineal generalizado de tipo logístico ordinal (22). Se seleccionó un término de error con una desviación estándar de 5 para la variable respuesta y una de partición del Propensity Score en 5 cuartiles, como lo sugiere la literatura (46).

En la literatura la mayoría de estudios existentes de esta clase evalúan el desempeño del Propensity Score comparado con la regresión lineal convencional o la regresión logística (6,37,48), mientras que el presente estudio no se centró solamente en dicha comparación, buscó determinar las condiciones necesarias para determinar que metodología seria optima ante un escenario específico de tamaño de muestra, relación expuesto: no expuesto y número de variables.

Dentro de los resultados observados se tienen que un mayor tamaño de muestra produce estimaciones más precisión independiente de las metodologías utilizadas (Regresión lineal convencional, ponderación media marginal basada en un Propensity Score mediante el uso de MMWS o el uso de estimadores doblemente robustos DR-MMWS). Está claro que estos estimadores producen estimados razonablemente insesgados del efecto de tratamiento verdadero a medida que el tamaño de muestra crece. Dichas estimaciones producen resultados similares, aunque al comparar los estimadores de regresión por mínimos cuadrados ordinarios y los estimadores que combinan Propensity Score con MMWS, los resultados se inclinan a favor de la regresión lineal convencional. Un aumento en la relación expuesto: no expuesto, aumentan la potencia tanto del modelo de regresión como el Propensity Score a través de MMWS, manteniéndose el número de variables

constante. Adicionalmente, el gasto computacional como la teoría envuelta en la implementación de un Propensity Score basado en MMWS es más costoso y engorroso en comparación con la regresión lineal, este último tendría un punto a favor de esta última metodología. Sin embargo, los computadores de la actualidad pueden tratar con estos algoritmos con relativa facilidad.

El modelo hace un supuesto sobre la relación entre el tratamiento y las variables modificadoras del efecto importante, el supuesto de riesgos proporcionales (16,22) para este estudio la simulación partió del hecho de que no se incumplía este supuesto, sin embargo, este hecho puede llegar a ser una limitación del presente estudio. La inclusión de un modelo logístico ordinal generalizado (22) se plantea como una solución a esta limitación, aunque computacionalmente implicaría un costo adicional en la estimación del efecto del tratamiento/intervención a través de la metodología de Propensity Score conjunto con MMWS. Otra limitación a tener en cuenta es que los datos observacionales en la investigación en salud suelen estar cargados de factores no observables que pueden afectar tanto los grados de exposición como del desenlace observados (37). Otro aspecto a tener en cuenta en el proceso de simulación es la independencia de las covariables simuladas, para posteriores estudios sería interesante considerar variables con algún grado de correlación. Por último, debe tenerse en cuenta que, aunque “no hay un método superior” en cuanto a emparejamiento se refiere y que “debe escogerse el método que tenga el mejor equilibrio y que aún cumpla con el objetivo analítico”, es importante evaluar el uso de los otros métodos “marching” distintos al emparejamiento por radio en las comparaciones realizadas en este estudio, con el fin de determinar si los parámetros evaluados presentan cambios importantes en cuanto a la calidad de la estimación y la reducción del sesgo y sí el uso de otro tipo de emparejamiento puede mejorar los resultados mencionados en este estudio.

Fong, et al. (49) en su estudio menciona que “a pesar de los avances en la generalización de los métodos de puntaje de propensión a los tratamientos no binarios, los investigadores aplicados a menudo dicotomizan las variables de tratamiento no binarias con el fin de utilizar los métodos de puntaje de propensión”. Esto es considerado una mala práctica en el sentido que se pierde información que puede afectar los resultados de un estudio. A pesar de que este estudio utiliza otra metodología para ajustar el efecto de una variable exposición o un tratamiento a una respuesta continua, la conclusión es válida, en el sentido que ya existen modelos que permiten lidiar con esta clase de inconvenientes (44), por lo que debe minimizarse esta práctica.

Un aspecto para destacar en esta investigación es la evaluación de tantos aspectos que pueden influir en la estimación del efecto de una intervención o tratamiento en un desenlace de tipo continuo, un aspecto a tener en cuenta que surge de este trabajo y es mencionado en trabajos similares (47) es la evaluación del desempeño de un modelo basado en Propensity Score en una investigación particular. Muchos de los estudios que aplican alguna de las metodologías comparadas en el presente estudio, no proporcionan información sobre las variables incluidas en el modelo y el porqué de la selección de un modelo en particular.

Linden (44) en su estudio menciona que “cuando tanto el puntaje de propensión como los modelos de resultado están mal especificados, MMWS y DR-MMWS superan sustancialmente a todos los demás estimadores”, Adicionalmente, Linden usa otras metodologías que permiten ajustar el efecto de confusión en estudios observacionales, mencionando que el MMWS supera a IPTW (Inverse Probability of Treatment Weights) cuando el modelo de puntaje de propensión está mal especificado, Hong (40) sugiere que dado IPTW se calcula como una función directa de la puntuación de propensión estimada; cuando el puntaje de propensión estimado se especifica mal, el IPTW se desviará sistemáticamente del peso verdadero”. En este aspecto Hong (39) menciona por ejemplo que el método IPTW es conocido por el sesgo y la imprecisión en la estimación, especialmente cuando los modelos de puntaje de propensión están mal especificados en sus formas funcionales (43). Este comportamiento mencionado, solo se presentó en los estimadores DR-MMWS, cuyo rendimiento fue equiparable a los estimadores dados por una regresión lineal múltiple en el presente estudio.

Este estudio resalta que con tamaños de muestra superiores a 500 individuos los estimadores MMWS y DR-MMWS se comportan de la misma manera que los estimadores basados en mínimos cuadrados ordinarios que arroja un modelo de regresión lineal múltiple. Un último aspecto que es importante acotar es que este estudio busca determinar la metodología que mejor controla el fenómeno de la confusión para diversos escenarios de tamaño de muestra y no pretende evaluar el efecto de la presencia de sesgos de selección en la implementación de las metodologías tratadas en este estudio, en este aspecto se parte del supuesto de ignorabilidad para realizar las comparaciones descritas en este manuscrito.

En resumen, es importante para las investigaciones de tipo clínica implementar este tipo de metodologías, ya que estos métodos mejoran el equilibrio de las covariables, reducen

el sesgo en las estimaciones del efecto del tratamiento (44) y las estimaciones obtenidas tiene un grado de precisión importante y surgen como un método sustituto en especial si los supuestos de la regresión lineal múltiple son quebrantados.

8. Conclusiones y recomendaciones

8.1 Conclusiones

Este trabajo se basa en una metodología relativamente reciente conocida como MMWS (Ponderación Media Marginal a través de la Estratificación; Marginal Mean Weighting through Stratification en inglés), método que busca eliminar el sesgo de selección asociado con un gran número de covariables igualando la composición de pretratamiento en un escenario donde la variable de exposición/tratamiento de estudio es de tipo ordinal. El análisis de simulación presentado en este trabajo asumiendo el cumplimiento del supuesto de ignorabilidad y el supuesto de riesgos proporcionales para la estimación del Propensity Score con una regresión logística ordinal y usando un emparejamiento por radio condujo a las siguientes conclusiones:

- Un mayor tamaño de muestra definitivamente otorga mayor precisión a las estimaciones realizadas con cualquiera de las metodologías mostradas en el siguiente trabajo (Regresión lineal convencional o la ponderación media marginal basada en un Propensity Score mediante el uso de MMWS). Está claro que ambas metodologías producen resultados cada vez más insesgados del efecto de tratamiento verdadero a medida que el tamaño de muestra crece.
- Las metodologías comparadas producen resultados similares, aunque, los gráficos estimados y las medidas de comparación calculadas, revelan pequeñas diferencias relativas en el rendimiento en favor de la regresión lineal múltiple para tamaños de muestra inferiores a 500, superado este umbral los estimadores basado en MMWS, especialmente los estimadores doblemente robustos equiparan las propiedades de los estimadores basados en mínimos cuadrados ordinarios.
- A mayor razón no expuesto: expuesto, la potencia tanto del modelo de regresión como el Propensity Score a través de MMWS es mayor.

- No hay una ganancia importante en la inclusión de más variables, en términos de reducción de la variabilidad o del sesgo, independiente de la metodología implementada. Esto va de la mano con el principio de parsimonia que rige a los modelos matemáticos, donde se busca explicar un fenómeno en términos de no redundancia de información. Esto deja una duda sobre las variables no observadas en el sentido que no necesariamente la inclusión de un mayor número de variables no observadas implica una mejora en la estimación del efecto de un tratamiento o de una intervención.
- El uso de DR-MMWS es una estrategia interesante a tener en cuenta en el análisis en estudios observacionales, ya que las estimaciones son consistentes y presentan una reducción del sesgo importante.
- Tanto el gasto computacional como la teoría envuelta en la implementación de un Propensity Score basado en MMWS es más costoso y engorroso en comparación con la regresión lineal, este último tendría un punto a favor de esta última metodología. Sin embargo, ya se cuentan con rutinas implementadas en varios paquetes estadísticos (Stata o R) que permiten usar esta última metodología fácilmente y los computadores de la actualidad pueden tratar con estos algoritmos con relativa facilidad.

8.2 Recomendaciones

Para futuras investigaciones sobre el tema se recomienda realizar un estudio que involucre la implementación de un Propensity Score basado en una regresión logística ordinal generalizada con el fin de establecer una metodología más robusta, en el sentido que el supuesto de riesgos proporcionales sería relevante y podría establecerse puntajes de propensión basados en variables de tipo ordinal que midan niveles de exposición o tratamiento más allá de una medición dicotómica de presencia o ausencia de una variable en salud de interés, así como el uso de otros métodos de emparejamiento distintos al emparejamiento por radio. Adicionalmente, y con base en los estudios referenciados en el siguiente manuscrito se recomienda dar a conocer estas metodologías a los investigadores que se encuentran realizando proyectos de tipo observacional con el fin de difundir estas nuevas metodologías que van en beneficio de conclusiones más fidedignas con respecto a los estudios observacionales que desarrollan.

9. Bibliografía

1. Velez, M; Egurrola JJF. Ronda clínica y epidemiológica Uso de la puntuación de propensión (propensity score) en estudios no experimentales. *Iatreia*. 2013;26(1):95–101.
2. Emanuel, EJ; Wendler, D; Grady C. What makes clinical research ethical? *JAMA - J Am Med Assoc*. 2000;283:2701–2711.
3. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *J Am Stat Assoc [Internet]*. 1996;91(434):444–55. Available from: <http://www.jstor.org/stable/2291629>
4. Linden A, Adams JL, Roberts N. Evaluating disease management programme effectiveness: an introduction to the regression discontinuity design. *J Eval Clin Pract*. 2006 Apr;12(2):124–31.
5. Filleron T, Kwiatowski F. Propensity score: A credible alternative to randomization? *Bull Cancer*. 2016 Jan;103(1):113–22.
6. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol*. 2003 Aug;158(3):280–7.
7. Sackett DL. Bias in analytic research. *J Chronic Dis [Internet]*. 1979;32(1):51–63. Available from: <http://www.sciencedirect.com/science/article/pii/0021968179900122>
8. Sainani KL. Propensity scores: uses and limitations. *PM R*. 2012 Sep;4(9):693–7.
9. Szklo, Moyses; Nieto J. *Epidemiología clínica: Conceptos y aplicaciones*. Díaz de lo. Silva, Luis; Jiménez R, editor. Madrid; 2003. 421 p.
10. Londoño JL. *Metodología de la investigación epidemiológica*. 6th ed. Moderno EEM, editor. Medellín; 2014. 365 p.
11. Gerstenblüth, M; Pagano J. Tratamiento de la endogeneidad y métodos de correspondencia en Stata. In 2008. p. 23.
12. Pattanayak, CW; Rubin, DB; Zell E. Métodos de puntuación de propensión para crear una distribución equilibrada de las covariables en los estudios observacionales. *Rev Española Cardiol*. 2011;64(10):897–903.
13. Patel VI, Conrad MF, Kwolek CJ, Ouriel K, Fairman RM, Cambria RP. Propensity-matched cohort validates findings of the VALOR trial. *J Vasc Surg*. 2011 Jul;54(1):22–9.
14. Biondi-Zoccai G, Romagnoli E, Agostoni P, Capodanno D, Castagno D, D’Ascenzo F, et al. Are propensity scores really superior to standard multivariable analysis? *Contemp Clin Trials*. 2011 Sep;32(5):731–40.
15. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for

- reducing model dependence in parametric causal inference. *Polit Anal* [Internet]. 2007 [cited 2018 Feb 11];15(3):199–236. Available from: <https://gking.harvard.edu/files/matchp.pdf>
16. Kereiakes DJ, Cannon LA, Ormiston JA, Turco MA, Mann T, Mishkel GJ, et al. Propensity-matched patient-level comparison of the TAXUS Liberte and TAXUS element (ION) paclitaxel-eluting stents. *Am J Cardiol*. 2011 Sep;108(6):828–37.
 17. Wang S, Meng W, An H, Zhao X, Lu J, Shi F. Propensity score matching-based therapeutic effect evaluation in patients with hepatocellular carcinoma. *J South Med Univ*. 2012 Sep;32(9):1234–7.
 18. Kojima S, Hiraki T, Gobara H, Iguchi T, Fujiwara H, Matsui Y, et al. Fracture of totally implanted central venous access devices: a propensity-score-matched comparison of risks for Groshong silicone versus polyurethane catheters. *J Vasc Access*. 2016 Nov;17(6):535–41.
 19. Boonstra PS, Bondarenko I, Park SK, Vokonas PS, Mukherjee B. Propensity score-based diagnostics for categorical response regression models. *Stat Med*. 2014 Feb;33(3):455–69.
 20. Boliver V, Swift A. Do comprehensive schools reduce social mobility? *Br J Sociol*. 2011 Mar;62(1):89–110.
 21. Zanutto E. A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data. Vol. 4, *Journal of Data Science*. 2006. 67-91 p.
 22. Linden A. Combining propensity score-based stratification and weighting to improve causal inference in the evaluation of health care interventions. *J Eval Clin Pract*. 2014 Dec;20(6):1065–71.
 23. Becker SO. *Methods to Estimate Causal Effects. Theory and Applications*. University of Stirling, editor. Stirling; 2009. 199 p.
 24. David, Kleinbaum; Lawrence, Kupper; Keith E MAN. *Applied Regression Analysis and Other Multivariable Methods*. 4th ed. Inc WPC, editor. 2007. 991 p.
 25. Kleinbaum, David G; Klein M. *Logistic Regression. A Self-Learning Text*. 3rd ed. Springer New York, editor. 2010. 702 p.
 26. McCullagh, Peter; Nelder J. *Generalized Linear Models*. 1989.
 27. Vargas JA. *Modelos Lineales Generalizados*. 1st ed. Chavéz P, editor. BOGOTÁ D. C.: Universidad Nacional de Colombia, Sede Bogotá; 2008. 105 p.
 28. Rosenbaum PR. *Observational Studies*. 2nd ed. Springer, editor. 2002. 375 p.
 29. Geoffrey, Montgomery; Douglas, Peck; Elizabeth V. *Introducción al análisis lineal*. Tercera ed. México: CECSA; 2006. 612 p.
 30. Chatterjee, Samprit; Hadi A. *Regression analysis by example*. Fourth edi. New Jersey; 2006. 383 p.

31. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis [Internet]. Springer; 2001 [cited 2014 May 5]. 568 p. Available from: <http://books.google.com/books?id=kfHrF-bVcvQC&pgis=1>
32. Long, J. Scott; Freese J. Regression models for categorical dependent variables using Stata. 2nd ed. Texas: Stata Corporation; 2001. 311 p.
33. Hollander M, Wolfe DA, Chicken E. Nonparametric statistical methods. Vol. 2, Nurse anesthesia. 2014. 93-95 p.
34. Gibbons, Jean Dickinson; Chakraborti S. Nonparametric Statistical Inference. 5th ed. Press C, editor. 2010. 650 p.
35. Gil-Flores J. Aplicación del método Bootstrap al contraste de hipótesis en la investigación educativa. Revista de educación, ISSN 0034-8082, N° 336, 2005, pags. 251-265. 2005.
36. J. Bickel P, M. Krieger A. Confidence Bands for a Distribution Function Using the Bootstrap. Vol. 84, Journal of The American Statistical Association - J AMER STATIST ASSN. 1989. 95-100 p.
37. Marie K-Mi. Nonparametric Statistics for Social and Behavioral Sciences. Press C, editor. 2013. 260 p.
38. Guo, Shenyang; Fraser MW, Guo S, Fraser MW. Propensity Score Analysis Statistical Methods and Applications. 2nd ed. Oaks T, editor. Sage Publications; 2015. 370 p.
39. Waernbaum I. Model misspecification and robustness in causal inference: comparing matching with doubly robust estimation. Stat Med. 2012 Jul;31(15):1572–81.
40. Feng P, Zhou X-H, Zou Q-M, Fan M-Y, Li X-S. Generalized propensity score for estimating the average treatment effect of multiple treatments. Stat Med. 2012 Mar;31(7):681–97.
41. Pijoán J, Martínez L, Arantza U. Ensayos clínicos vs estudios observacionales con índice de propersión para el estudio de asociaciones causales. In: ciberesp. Madrid; 2011. p. 1–40.
42. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. PLoS One. 2011 Mar;6(3):e18174.
43. Guo, Shenyang; Fraser MW, Guo S, Fraser MW, Austin PC, Small DS, Emsley R, et al. Propensity Score Analysis Statistical Methods and Applications [Internet]. 2nd ed. Oaks T, editor. Vol. 22, Statist. Sci. Texas, England: Sage Publications; 2015 [cited 2017 Nov 26]. 370 p. Available from: <http://www.jstor.org/stable/2291629>
44. Garrido MM, Kelley AS, Paris J, Roza K, Meier DE, Morrison RS, et al. Methods for constructing and assessing propensity scores. Health Serv Res. 2014 Oct;49(5):1701–20.
45. Wei P, Haiyan B. Propensity score analysis: Fundamentals and developments.

- New York, NY: The Guilford Press; 2015. 402 p.
46. Murnane, Richard J.; Willett JB. *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. OUP; 2010. 416 p.
 47. Dehejia RH, Wahba S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *J Am Stat Assoc* [Internet]. 1999;94(448):1053–62. Available from: <http://www.jstor.org/stable/2669919>
 48. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv* [Internet]. 2008;22. Available from: <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
 49. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat* [Internet]. 2011;10. Available from: <https://doi.org/10.1002/pst.433>
 50. Moreno-Serra R. Matching estimators of average treatment effects: a review applied to the evaluation of health care programmes. *Heal Econom Data Gr Work Pap* [Internet]. 2007 [cited 2018 May 26];(February). Available from: https://www.york.ac.uk/media/economics/documents/herc/wp/07_02.pdf
 51. Chen VW, Zeiser K. Implementing Propensity Score Matching Causal Analysis with Stata. Retrieved July [Internet]. 2008 [cited 2018 May 26];25(May):2010. Available from: <https://www.stata.com/meeting/7uk/sianesi.pdf>
 52. Zhang Z, Ni H, Xu X. Observational studies using propensity score analysis underestimated the effect sizes in critical care medicine. *J Clin Epidemiol*. 2014 Aug;67(8):932–9.
 53. Stemkowski, S., Saunders, W., & Yang G. *A Review of Propensity Score Application in Healthcare Outcome and Epidemiology*. 2007;
 54. Austin PC, Small DS. The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Stat Med* [Internet]. 2014 Oct 30;33(24):4306–19. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4260115/>
 55. Yu B, Gastwirth JL. A method of assessing the sensitivity of the Cochran-Mantel-Haenszel test to an unobserved confounder. *Philos Trans R Soc A Math Phys Eng Sci* [Internet]. 2008 [cited 2018 Feb 12];366(1874):2377–88. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2446605/pdf/rsta20080030.pdf>
 56. Johnson ML, Bush RL, Collins TC, Lin PH, Liles DR, Henderson WG, et al. Propensity score analysis in observational studies: outcomes after abdominal aortic aneurysm repair. *Am J Surg*. 2006 Sep;192(3):336–43.
 57. Hong G. *Causality in a Social World: Moderation, Mediation and Spill-over*. Wiley, editor. 2015. 444 p.
 58. Hong G. Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. *J Educ Behav Stat*. 2010;35(5):499–531.

59. Hong G. Marginal mean weighting through stratification: a generalized method for evaluating multivalued and multiple treatments with nonexperimental data. *Psychol Methods*. 2012 Mar;17(1):44–60.
60. VanderWeele TJ. Ignorability and stability assumptions in neighborhood effects research. *Stat Med*. 2008;27(11):1934–43.
61. Robins JM, Rotnitzky A, Zhao LP. Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *J Am Stat Assoc* [Internet]. 1994;89(427):846–66. Available from: <http://www.jstor.org/stable/2290910>
62. Kang JDY, Schafer JL. Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Stat Sci* [Internet]. 2007 [cited 2017 Nov 26];22(4):523–39. Available from: <https://projecteuclid.org:443/euclid.ss/1207580167>
63. Linden A. Improving causal inference with a doubly robust estimator that combines propensity score stratification and weighting. *J Eval Clin Pract*. 2017 Aug;23(4):697–702.
64. Emsley R, Lunt M, Pickles A, Dunn G. Implementing double-robust estimators of causal effects. *Stata J* [Internet]. 2008;8(3):334–53. Available from: <http://www.stata-journal.com/article.html?article=st0149>
65. Linden A. A comparison of approaches for stratifying on the propensity score to reduce bias. *J Eval Clin Pract*. 2017 Aug;23(4):690–6.
66. Simmering J. The Problem with Propensity Scores [Internet]. *R-bloggers*. 2015 [cited 2017 Nov 26]. Available from: <https://www.r-bloggers.com/the-problem-with-propensity-scores/>

Anexos

A. Aprobación Comité de Ética



Comité de Ética
Facultad de Medicina
Sede Bogotá

UNIVERSIDAD
NACIONAL
DE COLOMBIA

ACTA DE EVALUACIÓN: N°. 016-238-17
Fecha: 26 octubre de 2017

Nombre completo del proyecto: "ANÁLISIS COMPARATIVO ENTRE EL AJUSTE POR REGRESIÓN LINEAL MÚLTIPLE Y PROPENSITY SCORE ORDINAL BASADO EN PONDERACIÓN MEDIA MARGINAL POR ESTRATIFICACIÓN (MMWS) EN EL CONTROL DE LA CONFUSIÓN EN ESTUDIOS OBSERVACIONALES".

Versión número: 01

Sometido por: el estudiante Sergio Mauricio Moreno López

Dirigido por: el profesor Óscar Alexander Guevara Cruz

Presentado por: el profesor Ricardo Sánchez Pedraza, Coordinador Académico

Departamento o Sección: Maestría en Epidemiología Clínica de la Facultad de Medicina

Fecha en que fue sometido a consideración del Comité: 26 de octubre de 2017

EL COMITÉ DE ÉTICA DE INVESTIGACION DE LA FACULTAD DE MEDICINA. Se constituyó mediante la Resolución 152, (Acta No. 43 del 5 de diciembre de 1996) actualizado mediante resolución 008 (acta 03 de 27 de enero de 2011), de Consejo de Facultad el Comité de Ética de investigación, el cual está regido por la Resolución 008430 del 4 de octubre de 1993 del Ministerio de Salud de Colombia que estableció las normas científicas, técnicas y administrativas para la investigación en salud; los principios de la Asamblea Médica Mundial expuestos en su Declaración de Helsinki de 1964, última revisión del año 2000; y el código de regulaciones federales, título 45, parte 46, para la protección de los sujetos humanos, del departamento de salud y servicios humanos de los institutos Nacionales de Salud de los Estados Unidos (Junio 18 de 1991).

1. Sus miembros revisaron los siguientes documentos del presente proyecto:

- ✓ Carta de presentación del proyecto generada por la unidad básica o el departamento.
- ✓ Copia de la evaluación de los jurados o pares académicos que evaluaron y aprobaron el trabajo).
- ✓ Copia del proyecto completo de investigación,
- ✓ Dos resúmenes ejecutivos
- ✓ Dos copias del consentimiento informado (en español y cuando la investigación lo amerite).
- ✓ Hojas de vida resumidas de los investigadores y coinvestigadores del proyecto.
- ✓ Consideraciones éticas según resolución 8430 Ministerio de Salud.
- ✓ Resultados de evaluación por otros comités (si aplica).

2. El presente proyecto fue evaluado y aprobado por los siguientes miembros del Comité:

1	Amador Luis Roberto	Departamento de Patología
2	Duarte Gutiérrez Liz Marcela	Asesora Jurídica Facultad de Medicina
3	Guerrero Fonseca Carlos Arturo	Departamento de Ciencias Fisiológicas
4	Parra Pineda Mario Orlando	Departamento de Obstetricia y Ginecología

3. El Comité consideró que el presente estudio:

a. Es válido desde el punto vista ético. La investigación involucra un riesgo igual al promedio para los sujetos que participan en ella. La investigación se ajusta a los estándares de la buena práctica clínica.

SESQUICENTENARIO
UNIVERSIDAD NACIONAL DE COLOMBIA



[Página 1/1]
Elaboró Jeannette Pineda A.

Carrera 30 N°. 45-03
FACULTAD DE MEDICINA, Edificio 471 - 1 piso, Of.136
Conmutador: (57) (1) 316 5000 ext. 15167
Bogotá, Colombia
eticasalud_fmibog@unal.edu.co

Patrimonio de todos los colombianos



Sede Bogotá

b. El Comité considera que las medidas que están siendo tomadas para proteger a los sujetos humanos son adecuadas.

4. El Comité informará inmediatamente a las directivas institucionales:

- Todo desacato de los investigadores a las solicitudes del Comité.
- Cualquier suspensión o terminación de la aprobación por parte del Comité.

5. El Comité informará inmediatamente a las directivas, toda información que reciba acerca de:

- Lesiones o daños a sujetos humanos con motivo de su participación en la investigación Problemas imprevistos que involucren riesgos para los sujetos u otras personas.
- Cualquier cambio o modificación a este proyecto que haya sido revisado y aprobado por este comité

6. Cuando el proyecto sea aprobado, será por un periodo de un (1) año a partir de la fecha de aprobación.

7. El Investigador principal deberá:

- Informar de cualquier cambio que se proponga introducir en el proyecto. Estos cambios no podrán ejecutarse sin la aprobación previa del COMITÉ DE ÉTICA DE LA FACULTAD DE MEDICINA) excepto cuando sean necesarios para minimizar o suprimir un peligro inminente o un riesgo grave para los sujetos que participan en la investigación.
- Avisar de cualquier situación imprevista que se considere implica algún signo de riesgo para los sujetos o la comunidad o el medio en el cual se lleva a cabo el estudio.
- Informar de cualquier evento adverso serio de algún paciente, comunicando la situación al secretario y al presidente del Comité de Ética), de acuerdo con la normatividad que el INVIMA ha generado a este respecto.
- Poner en conocimiento del comité toda información nueva importante respecto al estudio, que pueda afectar la relación riesgo/beneficio de los sujetos participantes.
- Comunicar cualquier decisión tomada por otros comités con respecto a la investigación que se lleva a cabo.
- Informar de la terminación prematura o suspensión del proyecto explicando las causas o razones.
- Presentar a este comité un informe cuando haya transcurrido un año, contado a partir de la aprobación del proyecto. Los proyectos con duración mayor a un año, serán reevaluados a partir del informe de avance integrado.
- Todos los proyectos deben entregar al finalizar un informe final de cierre del estudio, este cierre puede ser el informe final en formato completo o en formato de resumen de cierre de estudio, firmado por el investigador responsable del estudio.

8. Observaciones:

El comité considera que el proyecto de investigación no presenta dilemas éticos por lo tanto emite **Concepto Aprobatorio.**

Nombre: **CARLOS ARTURO GUERRERO FONSECA**
Título: PhD Doctorado en Bioquímica, MSc. en Farmacología y MSc. en Genética Humana
Cargo: Presidente Comité de Ética



B. Implementación de la simulación en R

```

library(foreign)
library(ggplot2)
library(MASS)
library(Hmisc)
library(reshape2)
library(dummies)
library(glue)
library(dplyr)
library(foreign)
library(R.oo)
library(SDMTools)

#####
## lmSim Function          ##
#####

rm(list = ls())
lmSim <- function(n, k, sampleN, ratio) {
  # generate k covariates
  X <- cbind(replicate(k, rnorm(n)), replicate(1, 0))
  X[, (k-1)] <- as.factor(cut2(X[, (k-1)], g=2))
  X[, (k-1)] <- dummy(X[, (k-1)]), [2]
  X[, k] <- as.factor(cut2(X[, k], g=3))
  X[, (k+1)] <- dummy(X[, k]), [3]
  X[, k] <- dummy(X[, k]), [2]

  # some errors to the propensity score model
  pErr <- rnorm(n, 0, 3)
  pt <- exp(1 + X %*% rep(1, (k+1)) + pErr) / (1 + exp(1 + X %*% rep(1, (k+1)) +
pErr))
  # assign treatment status
  t <- rbinom(n, 2, pt)

  # generate response
  yErr <- rnorm(n, 0, 5)
  y <- 1 + X %*% rep(1, (k+1)) + t + yErr

  # put into big data frame
  data <- data.frame(y = y, t = factor(t))
  data <- cbind(data, X)
  names(data) <- c("y", "t", paste(rep("x", (k+1)), 1:(k+1), sep = ''))

  # generate sample from population
  treatedSample1 <- data[t == 0, ][sample(1:nrow(data[t == 0, ]), ratio*sampleN,
T), ]
  treatedSample2 <- data[t == 1, ][sample(1:nrow(data[t == 1, ]), sampleN, T), ]
  treatedSample3 <- data[t == 2, ][sample(1:nrow(data[t == 2, ]), sampleN, T), ]

  sampleData <- rbind(treatedSample1, treatedSample2, treatedSample3)

  # fit OLS model

```

```

    model <- summary(lm(y ~ ., data = sampleData))
    results <- c(model$coefficients[2, 1], model$coefficients[2, 2],
model$coefficients[2, 4], model$coefficients[3, 1], model$coefficients[3, 2],
model$coefficients[3, 4])
    results
}

###Start simulation

#t1, t2
results1<- data.frame(matrix(NA, nrow = 4000 * 3, ncol = 7))
names(results1)<- c("estimatedEffect1", "ratio", "n", "method", "se", "p", "sig")
results1$ratio <- rep(c(1, 2, 4), 4000)
results1$n <- rep(seq(100, 2000, length.out = 100), each = 12)
results1$method <- rep(c("MLR"), each = 12000)

results2<- data.frame(matrix(NA, nrow = 4000 * 3, ncol = 7))
names(results2)<- c("estimatedEffect2", "ratio", "n", "method", "se", "p", "sig")
results2$ratio <- rep(c(1, 2, 4), 4000)
results2$n <- rep(seq(100, 2000, length.out = 100), each = 12)
results2$method <- rep(c("MLR"), each = 12000)

set.seed(136)
k<- 6
for (i in 1:12000) {
results1$estimatedEffect1[i] <- c(as.numeric(lmSim(5000, k, results1$n[i],
results1$ratio[i])[1]))
results1$se[i] <- c(as.numeric(lmSim(5000, k, results1$n[i],
results1$ratio[i])[2]))
results1$p[i] <- c(as.numeric(lmSim(5000, k, results1$n[i],
results1$ratio[i])[3]))
results1$sig[i] <- results1$p[i] <= 0.05
results2$estimatedEffect2[i] <- c(as.numeric(lmSim(5000, k, results2$n[i],
results2$ratio[i])[4]))
results2$se[i] <- c(as.numeric(lmSim(5000, k, results2$n[i],
results2$ratio[i])[5]))
results2$p[i] <- c(as.numeric(lmSim(5000, k, results2$n[i],
results2$ratio[i])[6]))
results2$sig[i] <- results2$p[i] <= 0.05
}

#Save data

setwd()

datos1m1<- write.table(results1, "C:/Users/smmor/Dropbox/Caimed/Epidemiologia
Clinica/Tesis/Programacion/results1m1.txt", sep="\t")
datos1m2<- write.table(results2, "C:/Users/smmor/Dropbox/Caimed/Epidemiologia
Clinica/Tesis/Programacion/results1m2.txt", sep="\t")

datos1m1<- write.table(results1, "C:/Users/Sergio.Moreno/Dropbox/Epidemiología y
Bioestadística/Epidemiologia Clinica/Tesis/Programacion/results1m1.txt",
sep="\t")

```

```
datoslm2<- write.table(results2, "C:/Users/Sergio.Moreno/Dropbox/Epidemiología y
Bioestadística/Epidemiologia Clinica/Tesis/Programacion/resultslm2.txt",
sep="\t")
```

```
#####
## MMWS Function          ##
#####
```

```
rm(list = ls())
mmwspsSim <- function(n, k, sampleN, ratio) {
  # generate k covariates
  X <- cbind(replicate(k, rnorm(n)), replicate(1, 0))
  X[, (k-1)] <- as.factor(cut2(X[, (k-1)], g=2))
  X[, (k-1)] <- dummy(X[, (k-1)]), 2]
  X[, k] <- as.factor(cut2(X[, k], g=3))
  X[, (k+1)] <- dummy(X[, k]), 3]
  X[, k] <- dummy(X[, k]), 2]

  # some errors to the propensity score model
  pErr <- rnorm(n, 0, 3)
  pt <- exp(1 + X %*% rep(1, (k+1)) + pErr)/(1 + exp(1 + X %*% rep(1, (k+1)) +
pErr))
  # assign treatment status
  t <- rbinom(n, 2, pt)

  # generate response
  yErr <- rnorm(n, 0, 5)
  y <- 1 + X %*% rep(1, (k+1)) + t + yErr

  # put into big data frame
  data <- data.frame(y = y, t = factor(t))
  data <- cbind(data, dummy(t), X)
  names(data) <- c("y", "t", paste(rep("t", 3), 1:3, sep = ''), paste(rep("x",
(k+1)), 1:(k+1), sep = ''))

  # generate sample from population
  treatedSample1 <- data[t == 0, ][sample(1:nrow(data[t == 0, ]), ratio*sampleN,
T), ]
  treatedSample2 <- data[t == 1, ][sample(1:nrow(data[t == 1, ]), sampleN, T), ]
  treatedSample3 <- data[t == 2, ][sample(1:nrow(data[t == 2, ]), sampleN, T), ]

  sampleData <- rbind(treatedSample1, treatedSample2, treatedSample3)
  keepsps <- c("t", paste(rep("x", (k+1)), 1:(k+1), sep = ''))
  dataps <- sampleData[, keepsps, drop= F]

  # estimate ordinal propensity scores with mmws using all k X predictors
  ps <- polr(factor(t) ~ ., Hess=TRUE, data=dataps)

  # find untransformed fitted values
  P <- predict(ps, type = "p")
  LGP <- log((P)/(1-P))

  # find matches, require be within 20% of a sd of the untransformed pooled ps
```

```

SUPPORT4 <- ifelse(LGP[,1] < (min(tapply(LGP[,1], sampleData$t, max)) +
sd(LGP[,1])*0.2),
                 ifelse(LGP[,1] > (max(tapply(LGP[,1], sampleData$t, min)) -
sd(LGP[,1])*0.2), 1, 0)
                 ,0)

data <- subset(cbind(sampleData, LGP[,1], SUPPORT4), SUPPORT4 == 1)
names(data) <- c("y", "t", paste(rep("t", 3), 1:3, sep = ''), paste(rep("x",
(k+1)), 1:(k+1), sep = ''), "LGP1", "SUPPORT4")

n.strata <- 5

LGP_t <- cut(data$LGP1, quantile(data$LGP1, prob = seq(0, 1, 1/n.strata)),
include.lowest = TRUE, labels = FALSE)
table_t <- table(LGP_t, data$t)

MMWS <- NULL
MMWS2 <- NULL
pz1 <- NULL
pz2 <- NULL
pz3 <- NULL
pz <- NULL
nmz1 <- NULL
nmz2 <- NULL
nmz3 <- NULL
nmz <- NULL
total <- sum(data$SUPPORT4)

for (i in 1:total){
  pz1[i] <- sum(data$t1)/total
  pz2[i] <- sum(data$t2)/total
  pz3[i] <- sum(data$t3)/total
  for (j in 1:n.strata) {
    if (LGP_t[i] == j){
      nmz1[i] <- table_t[j, 1]
      nmz2[i] <- table_t[j, 2]
      nmz3[i] <- table_t[j, 3]
      nmz[i] <- table_t[j, 1]+table_t[j, 2]+table_t[j, 3]
      MMWS[i] <-
(data$t1[i]*((pz1[i]*(nmz[i])/nmz1[i])))+(data$t2[i]*((pz2[i]*(nmz[i])/nmz2[i]
)))+(data$t3[i]*((pz3[i]*(nmz[i])/nmz3[i]))))
    }
  }
}

data <- cbind(data, as.numeric(MMWS))
names(data) <- c("y", "t", paste(rep("t", 3), 1:3, sep = ''), paste(rep("x",
(k+1)), 1:(k+1), sep = ''), "LGP1", "SUPPORT4", "MMWS")

# fit OLS model with Ordinal Propensity Score

model <- summary(lm(y ~ factor(t), data = data, weights = data$MMWS))
results <- c(model$coefficients[2, 1], model$coefficients[2, 2],
model$coefficients[2, 4], model$coefficients[3, 1], model$coefficients[3, 2],
model$coefficients[3, 4])

```

```

results
}

#set.seed(136)
#mmwspsSim(2000, 10, 1000, 1)
#View(mmwspsSim(5000, 10, 100, 0.25))

###Start simulation

#t1 y t2
results1<- data.frame(matrix(NA, nrow = 4000 * 3, ncol = 7))
names(results1)<- c("estimatedEffect1", "ratio", "n", "method", "se", "p", "sig")
results1$ratio <- rep(c(1, 2, 4), 4000)
results1$n <- rep(seq(100, 2000, length.out = 100), each = 12)
results1$method <- rep(c("MMWS PS"), each = 12000)

results2<- data.frame(matrix(NA, nrow = 4000 * 3, ncol = 7))
names(results2)<- c("estimatedEffect2", "ratio", "n", "method", "se", "p", "sig")
results2$ratio <- rep(c(1, 2, 4), 4000)
results2$n <- rep(seq(100, 2000, length.out = 100), each = 12)
results2$method <- rep(c("MMWS PS"), each = 12000)

set.seed(136)
k<- 6
for (i in 1:(12000)) {
results1$estimatedEffect1[i] <- c(as.numeric(mmwspsSim(5000, k, results1$n[i],
results1$ratio[i])[1]))
results1$se[i] <- c(as.numeric(mmwspsSim(5000, k, results1$n[i],
results1$ratio[i])[2]))
results1$p[i] <- c(as.numeric(mmwspsSim(5000, k, results1$n[i],
results1$ratio[i])[3]))
results1$sig[i] <- results1$p[i] <= 0.05
results2$estimatedEffect2[i] <- c(as.numeric(mmwspsSim(5000, k, results2$n[i],
results2$ratio[i])[4]))
results2$se[i] <- c(as.numeric(mmwspsSim(5000, k, results2$n[i],
results2$ratio[i])[5]))
results2$p[i] <- c(as.numeric(mmwspsSim(5000, k, results2$n[i],
results2$ratio[i])[6]))
results2$sig[i] <- results2$p[i] <= 0.05
}

#Save data

datosmmws1<- write.table(results1, "C:/Users/smmor/Dropbox/Caimed/Epidemiologia
Clinica/Tesis/Programacion/resultsmmws1.txt", sep="\t")
datosmmws2<- write.table(results2, "C:/Users/smmor/Dropbox/Caimed/Epidemiologia
Clinica/Tesis/Programacion/resultsmmws2.txt", sep="\t")

#####
## DR-MMWS Function      ##
#####

rm(list = ls())
lmmwspsSim <- function(n, k, sampleN, ratio) {

```

```

# generate k covariates
X <- cbind(replicate(k, rnorm(n)), replicate(1, 0))
X[, (k-1)] <- as.factor(cut2(X[, (k-1)], g=2))
X[, (k-1)] <- dummy(X[, (k-1)]), 2]
X[, k] <- as.factor(cut2(X[, k], g=3))
X[, (k+1)] <- dummy(X[, k]), 3]
X[, k] <- dummy(X[, k]), 2]

# some errors to the propensity score model
pErr <- rnorm(n, 0, 3)
pt <- exp(1 + X %*% rep(1, (k+1)) + pErr)/(1 + exp(1 + X %*% rep(1, (k+1)) +
pErr))
# assign treatment status
t <- rbinom(n, 2, pt)

# generate response
yErr <- rnorm(n, 0, 5)
y <- 1 + X %*% rep(1, (k+1)) + t + yErr

# put into big data frame
data <- data.frame(y = y, t = factor(t))
data <- cbind(data, dummy(t), X)
names(data) <- c("y", "t", paste(rep("t", 3), 1:3, sep = ''), paste(rep("x",
(k+1)), 1:(k+1), sep = ''))

# generate sample from population
treatedSample1 <- data[t == 0, ][sample(1:nrow(data[t == 0, ]), ratio*sampleN,
T), ]
treatedSample2 <- data[t == 1, ][sample(1:nrow(data[t == 1, ]), sampleN, T), ]
treatedSample3 <- data[t == 2, ][sample(1:nrow(data[t == 2, ]), sampleN, T), ]

sampleData <- rbind(treatedSample1, treatedSample2, treatedSample3)
keepsps <- c("t", paste(rep("x", (k+1)), 1:(k+1), sep = ''))
dataps <- sampleData[ , keepsps, drop= F]

# estimate ordinal propensity scores with mmws using all k X predictors
ps <- polr(factor(t) ~ ., Hess=TRUE, data=dataps)

# find untransformed fitted values
P <- predict(ps, type = "p")
LGP <- log((P)/(1-P))

# find matches, require be within 20% of a sd of the untransformed pooled ps
SUPPORT4 <- ifelse(LGP[,1] < (min(tapply(LGP[,1], sampleData$t, max)) +
sd(LGP[,1])*0.2),
ifelse(LGP[,1] > (max(tapply(LGP[,1], sampleData$t, min)) -
sd(LGP[,1])*0.2), 1, 0)
,0)

data <- subset(cbind(sampleData, LGP[,1], SUPPORT4), SUPPORT4 == 1)
names(data) <- c("y", "t", paste(rep("t", 3), 1:3, sep = ''), paste(rep("x",
(k+1)), 1:(k+1), sep = ''), "LGP1", "SUPPORT4")

n.strata <- 5

```

```

LGP_t <- cut(data$LGP1, quantile(data$LGP1, prob = seq(0, 1, 1/n.strata)),
include.lowest = TRUE, labels = FALSE)
table_t <- table(LGP_t, data$t)

MMWS <- NULL
MMWS2 <- NULL
pz1 = NULL
pz2 = NULL
pz3 = NULL
pz <- NULL
nmz1 = NULL
nmz2 = NULL
nmz3 = NULL
nmz<- NULL
total <- sum(data$SUPPORT4)

for (i in 1:total){
  pz1[i] <- sum(data$t1)/total
  pz2[i] <- sum(data$t2)/total
  pz3[i] <- sum(data$t3)/total
  for (j in 1:n.strata) {
    if (LGP_t[i] == j){
      nmz1[i] <- table_t[j, 1]
      nmz2[i] <- table_t[j, 2]
      nmz3[i] <- table_t[j, 3]
      nmz[i] <- table_t[j, 1]+table_t[j, 2]+table_t[j, 3]
      MMWS[i] <-
(data$t1[i]*(((pz1[i]*(nmz[i]))/nmz1[i])))+(data$t2[i]*(((pz2[i]*(nmz[i]))/nmz2[i]
])))+(data$t3[i]*(((pz3[i]*(nmz[i]))/nmz3[i]))))
    }
  }
}

data <- cbind(data, as.numeric(MMWS))
names(data) <- c("y", "t", paste(rep("t", 3), 1:3, sep = ''), paste(rep("x",
(k+1)), 1:(k+1), sep = ''), "LGP1", "SUPPORT4", "MMWS")
keeps <- c("y", "t", paste(rep("x", (k+1)), 1:(k+1), sep = ''))
datareg <- data[ , keeps, drop= F]

# fit DR-MMWS: OLS model with Ordinal Propensity Score and full data

model <- summary(lm(y ~ ., data = datareg, weights = data$MMWS))
results <- c(model$coefficients[2, 1], model$coefficients[2, 2],
model$coefficients[2, 4], model$coefficients[3, 1], model$coefficients[3, 2],
model$coefficients[3, 4])
  results
}

###Start simulation

#t1 y t2
results1<- data.frame(matrix(NA, nrow = 4000 * 3, ncol = 7))
names(results1)<- c("estimatedEffect1", "ratio", "n", "method", "se", "p", "sig")
results1$ratio <- rep(c(1, 2, 4), 4000)

```

```

results1$n <- rep(seq(100, 2000, length.out = 100), each = 12)
results1$method <- rep(c("DR-MMWS"), each = 12000)

results2<- data.frame(matrix(NA, nrow = 4000 * 3, ncol = 7))
names(results2)<- c("estimatedEffect2", "ratio", "n", "method", "se", "p", "sig")
results2$ratio <- rep(c(1, 2, 4), 4000)
results2$n <- rep(seq(100, 2000, length.out = 100), each = 12)
results2$method <- rep(c("DR-MMWS"), each = 12000)

set.seed(136)
k<- 6
for (i in 1:(12000)) {
  results1$estimatedEffect1[i] <- c(as.numeric(lmmmwspsSim(5000, k, results1$n[i],
  results1$ratio[i])[1]))
  results1$se[i] <- c(as.numeric(lmmmwspsSim(5000, k, results1$n[i],
  results1$ratio[i])[2]))
  results1$p[i] <- c(as.numeric(lmmmwspsSim(5000, k, results1$n[i],
  results1$ratio[i])[3]))
  results1$sig[i] <- results1$p[i] <= 0.05
  results2$estimatedEffect2[i] <- c(as.numeric(lmmmwspsSim(5000, k, results2$n[i],
  results2$ratio[i])[4]))
  results2$se[i] <- c(as.numeric(lmmmwspsSim(5000, k, results2$n[i],
  results2$ratio[i])[5]))
  results2$p[i] <- c(as.numeric(lmmmwspsSim(5000, k, results2$n[i],
  results2$ratio[i])[6]))
  results2$sig[i] <- results2$p[i] <= 0.05
}

#Save data

datoslmmmws1<- write.table(results1,
"c:/Users/Sergio.Moreno/Documents/resultslmmmws1.txt", sep="\t")
datoslmmmws2<- write.table(results2,
"c:/Users/Sergio.Moreno/Documents/resultslmmmws2.txt", sep="\t")

```


C. Indicadores de desempeño para el tratamiento/exposición 1

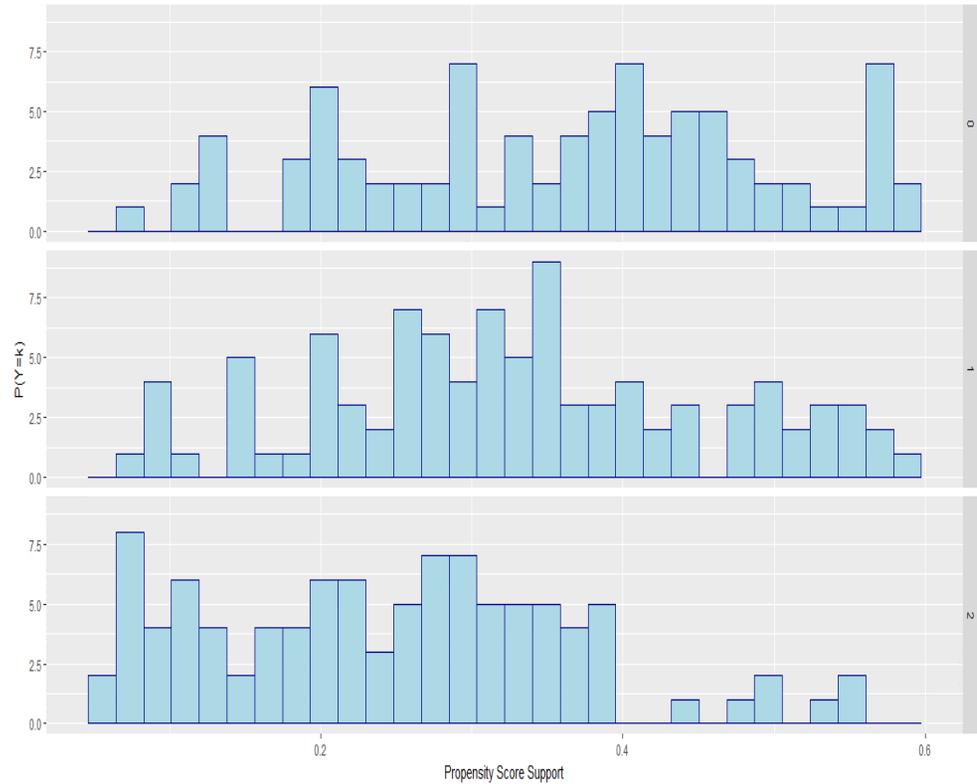
Razón no expuesto: expuesto	Tamaño de muestra	Regresión Lineal Múltiple (MLR)					Ponderación Media Marginal por Estratificación (MMWS)					Estimador doblemente robusto (DR-MMWS)				
		Efecto estimado	Error Estándar	MSE	MAPE	SMD	Efecto estimado	Error Estándar	MSE	MAPE	SMD	Efecto estimado	Error Estándar	MSE	MAPE	SMD
1:1	100-200	1,04	0,61	0,52	0,58	4,15%	1,08	0,65	0,66	0,66	8,46%	1,05	0,61	0,65	0,66	4,99%
	200-500	1,02	0,40	0,23	0,38	1,53%	1,08	0,42	0,28	0,42	8,43%	1,02	0,39	0,27	0,41	2,46%
	500-750	1,01	0,29	0,14	0,29	1,01%	1,07	0,31	0,18	0,33	7,32%	1,00	0,29	0,17	0,33	0,36%
	750-1000	0,99	0,25	0,11	0,26	-1,06%	1,06	0,26	0,13	0,29	6,17%	0,99	0,24	0,13	0,28	-1,25%
	1000-2000	1,00	0,19	0,09	0,24	-0,04%	1,08	0,20	0,11	0,27	8,01%	1,00	0,19	0,10	0,25	0,11%
1:2	100-200	1,04	0,53	0,30	0,44	3,93%	1,09	0,56	0,34	0,46	8,83%	1,04	0,52	0,33	0,45	3,94%
	200-500	1,02	0,34	0,18	0,33	2,45%	1,09	0,36	0,19	0,34	9,08%	1,03	0,34	0,19	0,33	2,73%
	500-750	0,99	0,25	0,12	0,28	-0,70%	1,06	0,27	0,14	0,29	6,29%	0,99	0,25	0,13	0,29	-0,54%
	750-1000	0,99	0,21	0,10	0,25	-1,34%	1,06	0,22	0,11	0,27	6,42%	0,99	0,21	0,11	0,26	-0,87%
	1000-2000	1,00	0,16	0,08	0,23	-0,22%	1,08	0,17	0,09	0,24	7,56%	1,00	0,16	0,09	0,24	-0,08%
1:4	100-200	1,03	0,48	0,31	0,44	2,75%	1,08	0,51	0,33	0,45	7,83%	1,03	0,48	0,32	0,45	3,26%
	200-500	0,99	0,31	0,16	0,33	-0,68%	1,05	0,33	0,17	0,33	5,29%	0,99	0,31	0,17	0,33	-0,66%
	500-750	1,03	0,23	0,11	0,26	2,87%	1,10	0,24	0,12	0,28	9,65%	1,02	0,23	0,11	0,26	2,45%
	750-1000	0,99	0,19	0,09	0,25	-0,55%	1,07	0,20	0,11	0,26	7,40%	1,00	0,19	0,10	0,25	-0,11%
	1000-2000	1,01	0,15	0,07	0,22	0,99%	1,09	0,16	0,09	0,23	8,50%	1,01	0,15	0,08	0,22	1,11%

D. Indicadores de desempeño para el tratamiento/exposición 2

Razón no expuesto: expuesto	Tamaño de muestra	Regresión Lineal Múltiple (MLR)					Ponderación Media Marginal por Estratificación (MMWS)					Estimador doblemente robusto (DR-MMWS)				
		Efecto estimado	Error Estándar	MSE	MAPE	SMD	Efecto estimado	Error Estándar	MSE	MAPE	SMD	Efecto estimado	Error Estándar	MSE	MAPE	SMD
1:1	100-200	2,03	0,67	0,42	0,25	1,34%	2,15	0,66	0,58	0,30	7,70%	2,03	0,62	0,55	0,29	1,60%
	200-500	2,00	0,43	0,23	0,19	0,03%	2,16	0,42	0,28	0,21	8,16%	2,01	0,40	0,26	0,21	0,60%
	500-750	1,99	0,32	0,14	0,15	-0,32%	2,17	0,31	0,19	0,17	8,55%	2,00	0,29	0,17	0,17	-0,01%
	750-1000	1,99	0,27	0,13	0,14	-0,66%	2,17	0,26	0,17	0,17	8,72%	1,99	0,24	0,14	0,15	-0,38%
	1000-2000	2,01	0,21	0,08	0,11	0,26%	2,20	0,20	0,13	0,15	9,75%	2,00	0,19	0,10	0,12	0,14%
1:2	100-200	1,94	0,58	0,38	0,23	-2,88%	2,08	0,57	0,46	0,26	4,09%	1,94	0,53	0,45	0,26	-2,80%
	200-500	1,97	0,37	0,19	0,17	-1,35%	2,14	0,36	0,24	0,19	7,07%	1,97	0,34	0,23	0,19	-1,55%
	500-750	1,97	0,28	0,11	0,14	-1,34%	2,16	0,27	0,17	0,16	7,80%	1,97	0,25	0,14	0,15	-1,46%
	750-1000	2,01	0,23	0,10	0,13	0,55%	2,20	0,23	0,15	0,16	10,09%	2,01	0,21	0,12	0,13	0,43%
	1000-2000	2,01	0,18	0,08	0,11	0,56%	2,21	0,17	0,13	0,14	10,63%	2,01	0,16	0,09	0,12	0,66%
1:4	100-200	1,97	0,51	0,30	0,22	-1,31%	2,09	0,51	0,46	0,27	4,57%	1,93	0,48	0,46	0,27	-3,31%
	200-500	2,02	0,33	0,17	0,17	1,01%	2,20	0,33	0,28	0,21	10,08%	2,02	0,31	0,24	0,20	1,10%
	500-750	2,01	0,25	0,10	0,13	0,73%	2,19	0,24	0,18	0,17	9,61%	2,00	0,23	0,14	0,15	-0,16%
	750-1000	2,00	0,21	0,10	0,12	-0,22%	2,21	0,21	0,15	0,16	10,33%	2,01	0,19	0,11	0,14	0,36%
	1000-2000	1,98	0,16	0,07	0,11	-0,86%	2,19	0,16	0,11	0,14	9,35%	1,98	0,15	0,08	0,11	-1,05%

E. Evaluación de soporte común y balance

n=300



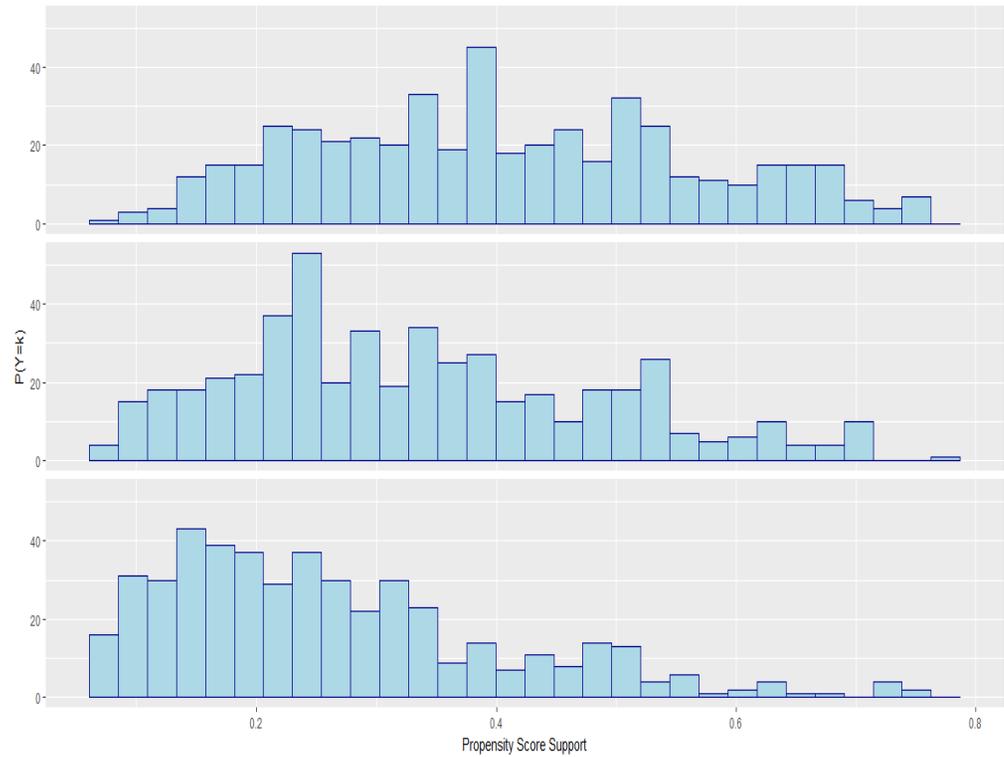
	T=0	T=1	T=2	Valor p
X1*	-0,087 (0,833)	-0,074 (1,02)	-0,199 (1,062)	0,628+
X2*	-0,178 (0,909)	-0,16 (0,858)	-0,089 (0,939)	0,771+
X3*	0,005 (1,02)	-0,135 (1,102)	0,075 (1,219)	0,419+
X4*	-0,178 (0,892)	-0,123 (1,093)	-0,119 (0,869)	0,898+
X5				
<i>dummy</i>	44 (50,57%)	43 (45,26%)	49 (49,49%)	0,746°
X6				
<i>nivel 2</i>	31 (35,63%)	39 (41,05%)	34 (34,34%)	0,895°
<i>nivel 3</i>	30 (34,48%)	30 (31,58%)	36 (36,36%)	

* Variables continuas resumidas en términos de promedios y D.E.

+Valor p de un ANOVA

°Valor p de una prueba Chi Cuadrado

n=1500



	T=0	T=1	T=2	Valor p
X1*	-0,164 (0,985)	-0,085 (0,903)	-0,163 (1,007)	0,340+
X2*	-0,216 (0,922)	-0,253 (0,992)	-0,196 (0,969)	0,644+
X3*	0,090 (0,938)	-0,03 (0,967)	-0,024 (0,978)	0,507+
X4*	-0,179 (0,972)	-0,081 (1,010)	-0,107 (0,866)	0,249+
X5				
dummy	44 (46,62%)	43 (40,84%)	49 (46,58%)	0,111*
X6				
nivel 2	31 (30,06%)	39 (33,20%)	34 (29,91%)	0,269*
nivel 3	30 (38,04%)	30 (40,64%)	36 (38,25%)	

* Variables continuas resumidas en términos de promedios y D.E.

+Valor p de un ANOVA

*Valor p de una prueba Chi Cuadrado