Imputación de datos faltantes: una aplicación del algoritmo de imputación multivariada por ecuaciones encadenadas (MICE) en salud pública

María Belén Arnaudo[1], María Soledad Fernández[1,2], and Adriana Alicia Pérez[1]

[1] Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina (GBA, FCEyN-UBA).
beluarnaudo@gmail.com
[2] Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina, Argentina (CONICET).

Resumen Los datos faltantes son muy comunes en las encuestas masivas, y se producen principalmente por falta de respuesta. Limitar el análisis a casos completos puede producir sesgos y pérdida de precisión en las estimaciones, pudiendo eventualmente debilitar la validez de los resultados y las conclusiones. La imputación múltiple mediante ecuaciones encadenadas constituye un enfoque flexible y práctico para manejar los datos faltantes. Este trabajo presenta una aplicación a partir del estudio del consumo de bebidas azucaradas en adolescentes y su asociación con determinantes sociales. Se utilizaron datos de la Encuesta Mundial de Salud Escolar (EMSE) 2012 y el paquete MICE de R. La muestra estuvo compuesta por 21107 adolescentes de 13 a 15 años pertenecientes a 561 escuelas de todo el país. La imputación múltiple permitió recuperar 6058 registros (28.7 % del total). Se encontró que el nivel educativo del hogar y de la escuela se asocian negativamente con el consumo de bebidas azucaradas: a menor nivel educativo, mayor riesgo de consumo. Siendo Argentina uno de los principales países consumidores de bebidas azucaradas del mundo, es fundamental que comiencen a desarrollarse estrategias para desincentivar este comportamiento, priorizando especialmente aquellos adolescentes pertenecientes a entornos de menores recursos.

Keywords: Datos faltantes · Encuesta Mundial de Salud Escolar · Argentina

1. Introducción

El sobrepeso y la obesidad en jóvenes y adolescentes están aumentando a nivel mundial, y la mayoría de los países de América Latina presenta la misma tendencia [22]. De acuerdo con los resultados de la última Encuesta Mundial de Salud Escolar (EMSE) [11] en Argentina 1 de cada 3 estudiantes de los primeros tres años del nivel medio presenta sobrepeso (39.1 % en varones vs. 27.8 % en mujeres), y 7.8 % presenta obesidad (10.6 % en varones y 5.3 % en mujeres). Factores como un mayor consumo y una mayor oferta de alimentos procesados

con alto contenido de sal, grasas y azúcares libres, bebidas azucaradas, un bajo consumo de frutas y una reducción progresiva de la actividad física, contribuyen al desarrollo de esta epidemia [21]. En particular, en Argentina la disponibilidad de consumo de gaseosas se duplicó entre 1996 y 2013 [31] y, si bien el consumo de bebidas azucaradas mostró una pequeña reducción entre las ediciones 2007 y 2012 de la EMSE [14], sigue siendo una fuente importante de calorías en adolescentes. De acuerdo a un estudio realizado en el 2020 por el Instituto de Efectividad Clínica y Sanitaria, el 71 % de los azúcares libres de los alimentos ultraprocesados son aportados por las bebidas azucaradas, y el 27 % de los casos de obesidad en la infancia es atribuible al consumo de bebidas azucaradas [1].

Para disminuir la prevalencia de obesidad y sobrepeso se debe no solo tener en cuenta las conductas saludables y no saludables individuales asociadas a ellos [18] como la dieta y la actividad física, sino también considerar sus determinantes sociales. Tales determinantes son múltiples y pertenecen a distintas escalas de organización que van desde el entorno inmediato (el hogar), hasta la escuela o el barrio [8]. Numerosos estudios en adolescentes han hallado relación entre el NSE del hogar y conductas vinculadas a la obesidad y sobrepeso, con mayor prevalencia de alimentación menos saludable en aquellos de menor NSE [16]. Existe evidencia que estas inequidades se están ampliando, en consonancia con la creciente desigualdad económica [10]. Teniendo en cuenta que los objetivos de desarrollo sustentable para el 2030 ponen especial foco en la equidad [27], y siendo Argentina, como el resto de los países de la región, un territorio con marcadas inequidades sociales, es de interés evaluar el impacto del nivel socioeconómico (NSE) a distintas escalas sobre las inequidades en salud.

Actualmente asistimos a una notable transformación en el abordaje de problemáticas vinculadas con la salud pública y las inequidades en salud en niños y adolescentes, propiciada por el acceso a encuestas con cobertura nacional y la incorporación de técnicas estadísticas más complejas. Una importante fuente de información de salud adolescentes en nuestro país es la Encuesta Mundial de la Salud Escolar. La población a la cual está destinada, la cobertura nacional y el volumen de datos que genera proporcionan características ideales para que la misma constituya una herramienta de gran utilidad para el estudio de la salud en adolescentes con un enfoque desde la ciencias de datos. Se trata de una encuesta estandarizada a nivel mundial elaborada por la Organización Mundial de la Salud (OMS) en colaboración con UNICEF, UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura), ONUSIDA (Programa Conjunto de las Naciones Unidas sobre el VIH/sida) y el CDC (Centros para el Control y Prevención de Enfermedades de Estados Unidos) que se aplicó en Argentina en los años 2007, 2012 y 2018. Tiene como objetivo relevar datos de los comportamientos y factores de riesgo y protección en adolescentes escolarizados de entre 13 a 15 años que se relacionan con las principales causas de enfermedad y muerte entre los jóvenes y adultos [11].

No obstante, es frecuente en este tipo de encuestas autoreportadas encontrar un gran número de datos faltantes [6]. La falta de respuesta puede causar problemas e introducir sesgos en las estimaciones si los datos faltantes se pro-

ducen por un mecanismo no completamente aleatorio, es decir, si quienes no contestaron a ciertas variables constituyen un grupo con algunas características específicas [25]. De no tratarse con técnicas específicas, los casos con datos faltantes son eliminados de los modelos de regresión (análisis de casos completos), lo que redunda en la eventual presencia de estimaciones sesgadas, menor precisión en las estimaciones y pérdida de potencia estadística en los análisis [17,24]. Especialmente grave es el problema cuando los adolescentes deben reportar el máximo nivel educativo alcanzado por cada uno de los padres [6]. Siendo esta la única variable asociada al NSE familiar en la EMSE, es evidente la necesidad de recurrir a métodos de imputación en los estudios sobre inequidades en salud adolescente.

Existen varios métodos propuestos en la literatura para el tratamiento de los datos faltantes, clasificados principalmente en métodos de imputación simple y métodos basados en modelos de distribución para los datos, como por ejemplo, la imputación múltiple [9]. Dentro de los métodos de imputación simple es frecuente el uso del reemplazo del valor faltante por la media o la moda, o imputación por regresión. Este tipo de soluciones simplistas muchas veces perjudica el análisis posterior de los datos, ya que resultan en al menos una subestimación sistemática de los errores estándar [5,9]. Por el contrario, la ventaja que presenta el uso de la imputación múltiple por sobre otras técnicas es que permite incorporar en las estimaciones la incertidumbre causada por los datos faltantes. Además, el método permite lidiar con casos más complejos tales como las variables derivadas, para las que es necesario mantener la coherencia al imputar (imputación pasiva), y la incorporación de términos de interacción en los modelos. Dentro de estos métodos se encuentra el método de Imputación Múltiple por ecuaciones encadenadas (MICE por sus siglas en inglés [28]). Este trabajo presenta una aplicación de este método con el paquete de R MICE al estudio de inequidades en salud, particularmente sobre el consumo de bebidas azucaradas en adolescentes a partir de la EMSE 2012. Se analizará cómo este hábito dietario no saludable -el consumo de bebidas azucaradas- es afectado por uno de los determinantes sociales en salud, como lo es el nivel socioeconómico, analizado a distintas escalas.

2. Materiales y Métodos

Fuente de datos

La EMSE 2012 en Argentina se implementó mediante un muestreo probabilístico de estudiantes de 1ro., 2do. y 3er. año de escuelas secundarias de todas las jurisdicciones del país (con excepción de Formosa). Se utilizó un muestreo por conglomerados bietápico. De un listado de escuelas públicas y privadas proporcionado por el Ministerio de Educación de la Nación, se seleccionaron 25 escuelas secundarias por jurisdicción, siendo que la probabilidad de selección de cada una fuera proporcional al número de estudiantes cursando 1ro a 3er año. Luego, se

seleccionaron al azar 2 o 3 cursos por escuela. El cuestionario fue autoadministrado y anónimo, e incluía una serie de módulos con un total de 81 preguntas que hacían referencia a diversos comportamientos vinculados a la salud adolescente. El trabajo de campo se llevó a cabo entre los meses noviembre y diciembre. El formulario e información detallada de la encuesta se encuentran disponibles en la página de la WHO.

Variables involucradas

Variable Respuesta. La pregunta a partir de la cual se evaluó el consumo de bebidas azucaradas inspeccionaba la frecuencia diaria de consumo en el último mes, siendo las opciones de respuesta: 'No tomé gaseosa en los últimos treinta días', 'Menos de una vez al día', 'una vez al día', 'dos veces al día', 'tres veces al día', 'cuatro veces al día' o 'cinco o más veces al día'. A partir de ella, se realizó una dicotomización para evaluar el consumo de al menos una bebida azucarada por día en el último mes (Si/No) y se la denominó, **'Consumo de Bebidas Azucaradas'**.

Variables Predictoras. Como variable indicadora de NSE del hogar se empleó el nivel educativo (NE) parental, que, si bien refleja solo una de las dimensiones del NSE, ya ha sido utilizado en estudios previos sobre obesidad y sobrepeso y conductas saludables y no saludables en adolescentes [16]. Para el NE a escala del hogar ('NE Hogar') se utilizó el nivel educativo máximo alcanzado por cualquiera de los padres/tutores, que se categorizó como 'Bajo' (primario incompleto o completo), 'Medio' (secundario incompleto o completo) y 'Alto' (terciario/universitario incompleto o completo). Para el NE a escala de la escuela ('NE Escuela') en primer lugar para cada estudiante se promedió el NE máximo de ambos padres/tutores, expresado en una escala de 1 (Primario Incompleto) a 6 (Terciario/Universitario Completo). Para cada escuela se agregaron los resultados del promedio obtenido para sus estudiantes, y se las clasificó de acuerdo a los terciles en NE "Bajo", "Medio" o "Alto". Como variables de control se incluyeron género, edad (considerada categórica con tres niveles, 13, 14 y 15 años) y porcentaje de hogares con necesidades básicas insatisfechas (NBI, Censo 2010 [20]) de cada provincia, de manera de ajustar por NSE jurisdiccional.

Proceso de imputación

En primer lugar se efectuó el diagnóstico de magnitud y distribución de los casos faltantes en las variables de interés. A continuación, se llevó a cabo una imputación múltiple de datos faltantes utilizando el algoritmo de imputación multivariada por ecuaciones encadenadas (MICE, Mice: Multivariate Imputation By Chained Equations) [4], que se basa en especificaciones completamente condicionales donde cada variable con datos faltantes es imputada en un modelo separado. Todas las variables de interés con datos faltantes fueron imputadas. Para realizar la imputación, se generó una matriz indicando las variables a imputar, cuáles se utilizarían como predictoras, y el método de imputación de acuerdo a la naturaleza de la variable. En este caso, como predictoras se usaron 15 variables

sociodemográficas y conductuales que formaban parte de la misma encuesta (año escolar, educación del padre y de la madre, veces que pasó hambre en la última semana, consumo de frutas, verduras, comida rápida, nivel de actividad física, utilización de bicicleta, horas de sedentarismo, víctima de bullying, motivo de bullying, pensamientos suicidas). Cabe destacar que, mediante la especificación del método, se pueden realizar imputaciones pasivas para mantener las relaciones entre variables y así, si una variable deriva de cálculos realizados sobre otra variable, la imputación pasiva mantendrá la coherencia entre ellas. En este caso, la variable bajo estudio, el consumo de bebidas azucaradas, es un ejemplo de variable derivada ya que proviene de la dicotomización de otra variable, y por ello se realizó una imputación pasiva. El nivel educativo máximo del hogar y el promedio agregado del nivel educativo por escuela también fueron imputados de esta forma.

El mecanismo consiste en tres pasos: imputación, análisis y agregado.

1. Imputación. Se generaron 30 bases imputadas, de acuerdo a las recomendaciones de Bodner (2008) [3] y White et al. (2011) [30]. En cada una de ellas se reemplazaron los valores faltantes por valores plausibles obtenidos de una distribución que varía según la escala de la variable. Cada base difería únicamente en los datos faltantes (imputados), y la magnitud de estas diferencias reflejaba la incerteza del valor imputado.
2. Análisis. El segundo paso consistió en implementar los modelos de interés en cada una de las bases generadas. Los estimadores de cada base serían diferentes a causa de la diferencia en la imputación.
3. Agregado. En el último paso se agregaron todos los estimadores en un único estimador final y se estimó su varianza, que contempla la varianza obtenida en cada imputación y la varianza entre imputaciones.

Para algunas bases imputadas no se logró convergencia del modelo, por lo que fueron descartadas del análisis y por ende del agregado.

Modelado de consumo de bebidas azucaradas

Con el objetivo de modelar el consumo de bebidas azucaradas según el nivel socioeconómico, se implementaron modelos lineales generalizados mixtos con distribución Bernoulli y función de enlace logit. Se utilizó la función glmer, del paquete lme4 [2].

El modelo incluyó la interacción entre NE escuela y NE hogar. Se incluyó a la escuela como factor aleatorio. Las estimaciones fueron por máxima verosimilitud mediante la cuadratura adaptativa de Gauss Hermite. En este caso, se trabajó con 5 puntos de cuadratura, lo que produce mejores estimaciones que trabajando con tan solo un punto. Se incluyó el optimizador BOBYQA ("Bound Optimization BY Quadratic Approximation") para aumentar la tasa de convergencia de los modelos [19]. El modelo propuesto fue:

$$
\begin{aligned}
logit(\pi_{i/j}) = \beta_0 &+ \beta_1 * Mujer_i + \beta_2 * Edad14_i + \beta_3 * Edad15_i + \beta_4 * NEHogMedio_i \\
&+ \beta_5 * NEHogAlto_i + \beta_6 * NEEscMedio_i + \beta_7 * NEEscAlto_i \\
&+ \beta_8 * NEHogMedio * NEEscMedio_i + \beta_9 * NEHogMedio * NEEscAlto_i \\
&+ \beta_{10} * NEHogAlto * NEEscMedio_i + \beta_{11} * NEHogAlto * NEEscAlto_i \\
&+ \beta_{12} * NBI\ jur_i + \alpha_j
\end{aligned}
$$

Donde:

i= 1 a 21107; *j: 1 a 561*

$Y_{ij} \sim Bernoulli(\pi_{i/j})$; $\alpha_j \sim N(0, \sigma^2_{Escuelas})$

Hog: Hogar; Esc: Escuela; jur: jurisdiccional

Se estimaron los odds ratio con sus respectivos intervalos de confianza al 95 %. A fin de evaluar su significancia se realizó el test de Wald comparando modelos anidados y se utilizó la función `pool.compare` del paquete MICE, que elimina términos del modelo respetando el principio de marginalidad.

Posteriormente, se analizó la existencia de diferencias en el consumo de bebidas azucaradas según el NE escuela dentro de cada categoría de NE a escala hogar. Para ello fue necesaria su implementación manual a partir de las bases individuales imputadas, ya que MICE no es compatible con librerías que efectúan dichas comparaciones, como emmeans.

Para estimar la proporción de varianza explicada por variaciones entre las escuelas, se calculó el coeficiente de correlación intraclase de cada base y luego se obtuvo el promedio del mismo para todas las bases.

3. Resultados

Fueron incluidos en los análisis 21107 estudiantes de entre 13 y 15 años pertenecientes a 561 escuelas, con una edad promedio de 14.4 años, 52.3 % mujeres. El 65 % de los encuestados reportó haber consumido bebidas azucaradas al menos una vez al día (Cuadro 1).

Las variables que mayor porcentaje de datos faltantes evidenciaron fueron el nivel educativo del padre y de la madre (20.8 % y 17.5 % respectivamente), variables que fueron los insumos para el cálculo del NE Hogar y NE Escuela. Tanto la variable respuesta como las variables de control género y edad no evidenciaron gran proporción de datos faltantes (1.01 %, 1.25 % y 1.11 % respectivamente). La variable NBI jurisdiccional (CENSO 2010) no presentó datos faltantes. En caso

de no haberse realizado la imputación, un análisis de casos completos hubiese resultado en una pérdida del 28.7 % de los casos. Se observa que los casos incompletos son estudiantes de menor edad, de NE del hogar bajo y mayor consumo de bebidas azucaradas, mientras que no se detectan diferencias en la proporción de géneros (Cuadro 1).

Cuadro 1. Comparación de casos completos versus casos incompletos -en al menos una variable- en relación a la edad, el género, el NE del Hogar y el consumo de bebidas azucaradas. Además, se muestra una descriptiva de la totalidad de los casos (Total).

Variable	**Casos completos** N=15480	**Casos incompletos** N= 5627	**p**	**Total**
Edad (años)			$< 0{,}01$	
13	3640 (23.5 %)	1347 (25.0 %)		4987 (23.9 %)
14	5991 (38.7 %)	2178 (40.4 %)		8169 (39.1 %)
15	5849 (37.8 %)	1868 (34.6 %)		7717 (37.0 %)
Género			0.139	
Mujer	8194 (52.9 %)	2902 (54.1 %)		11096 (53.2 %)
NE_Hogar			$< 0{,}01$	
Bajo	2670 (17.2 %)	102 (25.4 %)		2772 (17.5 %)
Medio	7394 (47.8 %)	187 (46.5 %)		7581 (47.7 %)
Alto	5416 (35.0 %)	113 (28.1 %)		5529 (34.8 %)
Bebidas Azucaradas			$< 0{,}01$	
Consume	9913 (64.0 %)	3733 (69.0 %)		3646 (65.3 %)

Se estimó el modelo para predecir el consumo de bebidas azucaradas en adolescentes en función de las variables demográficas y socioeconómicas a partir de 15 bases imputadas que convergieron. La interacción entre NE Escuela y NE Hogar resultó no significativa (p=0.31), por lo que se muestra el modelo sin interacción (Cuadro 2). No se observaron diferencias significativas entre los géneros (p=0.44). Se encontró que el NE del hogar y el de la escuela están asociados negativamente con el consumo de bebidas azucaradas, ajustando por NBI juridisccional. Es decir que los estudiantes pertenecientes a hogares de menor NE y a escuelas de menor NE poseen mayor riesgo de consumo que los pertenecientes a hogares de mayor NE o que asisten a escuelas de mayor NE (Figura 1). Menos del 20 % de la varianza aleatoria está explicada por la variación entre las escuelas (CCI = 0,19), indicando que hay más variación dentro de las escuelas que entre ellas.

Cuadro 2. Resultados del modelo aditivo para consumo de bebidas azucaradas en adolescentes escolarizados (15 imputaciones).

Variable		OR	IC_{95}	**p-valor**
Personales				
Edad (años)				$< 0{,}01$
	13	1		
	14	1.22	[1.19 ; 1.25]	
	15	0.94	[0.92 ; 0.96]	
Género				0.44
	Varón	1		
	Mujer	0.99	[0.98 ; 1.01]	
Sociodemográficas				
NE Hogar				$< 0{,}01$
	Bajo	1		
	Medio	0.98	[0.92 ; 1.05]	
	Alto	0.75	[0.70 ; 0.81]	
NE Escuela				$< 0{,}05$
	Bajo	1		
	Medio	1.04	[0.86 ; 1.24]	
	Alto	0.78	[0.65 ; 0.94]	
NBI jurisdiccional	-	1.03	[1.01 ; 1.05]	$< 0{,}01$

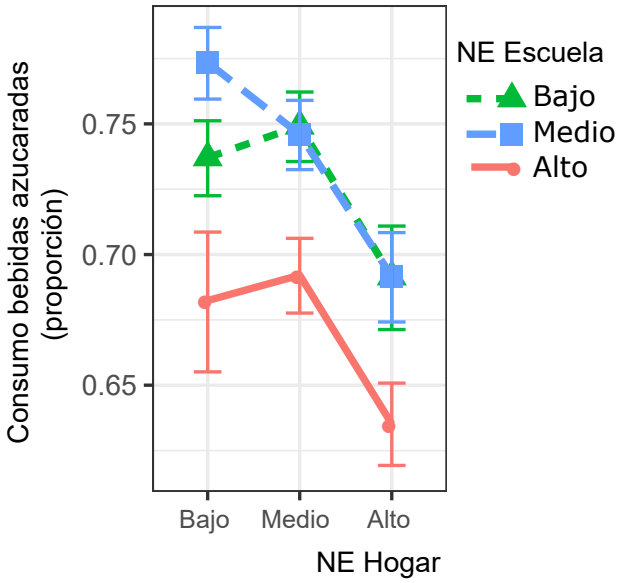

Figura 1. Proporción estimada de adolescentes escolarizados que consumen bebidas azucaradas al menos una vez por día. Los datos se muestran para la edad de 14 años.

4. Discusión

Este trabajo constituye un ejemplo de aplicación de un método de imputación aplicado a un problema en salud pública, como es el estudio de las inequidades en salud en adolescentes. La ausencia de respuesta fue muy común en la EMSE; de no haberse aplicado un método de imputación y efectuando un análisis de casos completos, se hubieran perdido 6058 registros (28.7 % del total). La omisión de registros con casos incompletos se asocia a sesgos y pérdida de precisión en las estimaciones, sobre todo si el mecanismo de generación de datos faltantes no es completamente al azar (MCA, missing completely at random) [23,15]. En este caso, se detectaron diferencias en múltiples variables entre los estudiantes que tenían completas todas las respuestas y los que no, sugiriendo una pérdida no completamente aleatoria y la eventual aparición de sesgos en el caso de analizarse solo casos completos.

La imputación múltiple por ecuaciones encadenadas constituye una herramienta eficiente y sumamente conveniente para lidiar con los problemas derivados de la no respuesta. Además de proporcionar estimaciones no sesgadas de los parámetros, incorpora la variación aleatoria entre las múltiples bases imputadas de manera de obtener buenas estimaciones de los errores estándar de los estimadores y puede ser aplicada a cualquier tipo de variables. Tiene como supuesto que el mecanismo de generación de los casos faltantes debe ser aleatorio (MAR, missing at random). Es decir, que la pérdida se asocia a una variable que ha sido registrada y no al valor no registrado, en cuyo caso el mecanismo sería no aleatorio (MNAR, missing not at random) [23]. Si bien con la información disponible no fue posible descartar que las pérdidas observadas en la EMSE fuesen no aleatorias, ha sido reportado que la imputación múltiple es razonablemente robusta frente a violaciones moderadas de este supuesto [12].

A pesar de que se han mencionado numerosos beneficios de trabajar con la imputación múltiple, es también pertinente señalar limitaciones de la herramienta. Por un lado, presenta un requerimiento computacional elevado, y por ello puede ser demandante en tiempo y recursos. Además, existen dificultades para incorporar al análisis el diseño muestral, estructuras complejas de anidamiento o datos longitudinales. No obstante, la imputación múltiple es tan solo una de las numerosas técnicas disponibles para lidiar con los datos faltantes y, por lo tanto, resultaría de interés para futuros trabajos repetir el análisis utilizando otra de las opciones analíticas y así poder comparar las metodologías. Particularmente, se podría recurrir a un método de imputación iterativa basado en random forest, que presenta la ventaja de ser menos costoso computacionalmente y, tal como MICE pero a diferencia de otros métodos, posibilita trabajar con variables de distinta naturaleza manteniendo las relaciones entre ellas [26].

Con respecto al análisis de los factores asociados al consumo de bebidas azucaradas en adolescentes, se encontró una asociación inversa con el NE del hogar: a mayor NE del hogar disminuye el consumo. En un estudio involucrando 28

países Vereecken et al., [29] encontraron asociación inversa entre el consumo de bebidas azucaradas y la escala de afluencia familiar en países europeos de mayores ingresos (Europa del norte, oeste y sur), mientras que en Europa central y del este, la asociación fue directa. Los autores argumentan que este comportamiento contrastante podría atribuirse al alto costo de las bebidas azucaradas en los países de Europa central y del este, lo que las haría únicamente accesibles para grupos de alto poder adquisitivo. En cambio, para el resto de los países donde el costo no es prohibitivo, podría especularse que en familias de mayor NE hay mayor conocimiento del efecto perjudicial de este tipo de bebidas, y que existen otros mecanismos protectores, tales como una mejor comunicación con los padres [7], que pueden prevenir comportamientos de riesgo y favorecer la adopción de comportamientos saludables. En nuestro país, el costo de las bebidas azucaradas es relativamente bajo [13], por lo que nuestros resultados son consistentes con los hallados por estos autores. Otro de nuestros resultados da cuenta del gradiente en el consumo según NE de la escuela. En las escuelas de mayores recursos hay menor consumo de bebidas azucaradas, lo que implica que el NE alto de la escuela podría tener un efecto protector en la salud de los adolescentes. Esto resalta la importancia de propiciar entornos saludables en las escuelas. En 2012, la EMSE reveló que solamente el 5.6 % de las escuelas observadas contaban con bebederos en los patios o dispensers de agua, mientras que el 80.2 % de las escuelas contaban con al menos un kiosco, de los cuales el 91.4 % ofrecía productos que por sus características nutricionales no están recomendados en la población adolescente. La presencia de un entorno obesogénico, donde los adolescentes tienen alta exposición y fácil acceso y disponibilidad de estos productos, incentiva la elección de alimentos no saludables, favoreciendo la mala nutrición y el sobrepeso.

Argentina es uno de los mayores consumidores de bebidas azucaradas del mundo y sería necesario que se implementen políticas públicas efectivas y basadas en evidencia para desalentar este consumo y proteger la salud, sobre todo de los grupos más vulnerables. Nuestros resultados, más robustos metodológicamente gracias a la implementación de imputación múltiple, aportan evidencia acerca de la elevada prevalencia de consumo de bebidas azucaradas en adolescentes escolarizados y de la brecha existente según el NE de los hogares y las escuelas. En ese sentido, estrategias impulsadas desde el estado, como los entornos saludables en las escuelas, la ley de etiquetado frontal y la implementación impuestos políticas de impuestos al consumo de bebidas azucaradas [13] pueden contribuir al reducir el consumo y mejorar la salud de los adolescentes.

Referencias

1. Alcaraz, A., Bardach, A., Espinola, N., Perelli, L., Balan, D., Cairoli, F., Palacios, A., Comolli, M., Augustovski, F., Pichon-Riviere, A.: El lado amargo de las bebidas azucaradas en Argentina. Tech. rep., Instituto de Efectividad Clínica y Sanitaria, Buenos Aires (2020)
2. Bates, D., Maechler, M., Bolker, B., Walker, S., Others: lme4: Linear mixed-effects models using Eigen and S4. R package version **1**(7), 1–23 (2014)
3. Bodner, T.E.: What improves with increased missing data imputations? Structural Equation Modeling: A Multidisciplinary Journal **15**(4), 651–675 (2008)
4. van Buuren, S., Groothuis-Oudshoorn, K.: mice: Multivariate imputation by chained equations in R. Journal of statistical software pp. 1–68 (2010)
5. Calafati, R.O.: Estrategias para el tratamiento de datos faltantes ("missing data") en estudios con datos longitudinales. Ph.D. thesis, Universitat Oberta de Catalunya (2017)
6. Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., Richter, M.: Researching health inequalities in adolescents: the development of the health behaviour in school-aged children (hbsc) family affluence scale. Social science & medicine **66**(6), 1429–1436 (2008)
7. Currie, C., Zanotti, C., Morgan, A., Currie, D.: Social determinants of health and well-being among young people. HBSC international report from the 2009/2010 Survey: Social determinants of health and well-being among young people. HBSC international report from the 2009/2010 Survey. WHO Regional Office for Europe (2012)
8. Diez Roux, A.V.: The Study of Group-Level Factors in Epidemiology : Rethinking Variables , Study Designs , and Analytical Approaches **26**, 104–111 (2004). https://doi.org/10.1093/epirev/mxh006
9. Donders, A.R.T., Van Der Heijden, G.J., Stijnen, T., Moons, K.G.: A gentle introduction to imputation of missing values. Journal of clinical epidemiology **59**(10), 1087–1091 (2006)
10. Elgar, F.J., Pförtner, T.K., Moor, I., De Clercq, B., Stevens, G.W., Currie, C.: Socioeconomic inequalities in adolescent health 2002–2010: a time-series analysis of 34 countries participating in the health behaviour in school-aged children study. The Lancet **385**(9982), 2088–2095 (2015)
11. EMSE: Encuesta Mundial de Salud Escolar Argentina 2012. (Septiembre), 50 (2014)
12. Faris, P.D., Ghali, W.A., Brant, R., Norris, C.M., Galbraith, P.D., Knudtson, M.L., Investigators, A., et al.: Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. Journal of clinical epidemiology **55**(2), 184–191 (2002)
13. Fernández, A., Mejía, R.M.: B.A.S.T.A. Bebidas Azucaradas, Salud y Tarifas en Argentina. Enfoque Multidisciplinario. CEDES (2018)
14. Ferrante, D., Linetzky, B., Ponce, M., Goldberg, L., Konfino, J., Laspiur, S.: Prevalencia de sobrepeso, obesidad, actividad fisica y tabaquismo en adolescentes argentinos: Encuestas Mundiales de Salud Escolar y de Tabaco en Jóvenes, 2007-2012. Archivos argentinos de pediatria **112**(6), 500–504 (2014)
15. Graham, J.W.: Missing data analysis: Making it work in the real world. Annual review of psychology **60**, 549–576 (2009)
16. Hanson, M.D., Chen, E.: Socioeconomic status and health behaviors in adolescence: a review of the literature. Journal of behavioral medicine **30**(3), 263 (2007)

17. Hendry, G.M., Naidoo, R.N., Zewotir, T., North, D., Mentz, G.: Model development including interactions with multiple imputed data. BMC medical research methodology **14**(1), 136 (2014)
18. Organización Mundial de la Salud: Enfoques poblacionales de la prevención de la obesidad infantil. Tech. rep., Ginebra (2016)
19. Powell, M.J.D.: The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge pp. 26–46 (2009)
20. Redatam, B.D.: Censo Nacional de Población , Hogares y Viviendas 2010 Censo del Bicentenario (2013)
21. Rennie, K.L., Johnson, L., Jebb, S.A.: Behavioural determinants of obesity. Best Practice & Research Clinical Endocrinology & Metabolism **19**(3), 343–358 (2005)
22. Rivera, J.Á., González de Cossío, T., Pedraza, L.S., Aburto, T.C., Sánchez, T.G., Martorell, R.: Childhood and adolescent overweight and obesity in Latin America: a systematic review. The lancet Diabetes & endocrinology **2**(4), 321–332 (2014)
23. Rubin, D.B.: Inference and missing data. Biometrika **63**(3), 581–592 (1976)
24. Rubin, D.B., Schenker, N.: Multiple imputation in health-are databases: An overview and some applications. Statistics in medicine **10**(4), 585–598 (1991)
25. Shavers, V.L.: Measurement of Socioeconomic Status in Health Disparities Research **99**(9), 1013–1023 (2007)
26. Stekhoven, D.J., Bühlmann, P.: Missforest—non-parametric missing value imputation for mixed-type data. Bioinformatics **28**(1), 112–118 (2012)
27. UN: Transforming our world: the 2030 agenda for sustainable development. Tech. rep., United Nations (2015)
28. van Buuren, S., Groothuis-Oudshoorn, K.: mice: Multivariate imputation by chained equations in r. Journal of Statistical Software **45**(3), 1–67 (2011), `https://www.jstatsoft.org/v45/i03/`
29. Vereecken, C.A., Inchley, J., Subramanian, S., Hublet, A., Maes, L.: The relative influence of individual and contextual socio-economic status on consumption of fruit and soft drinks among adolescents in europe. The European Journal of Public Health **15**(3), 224–232 (2005)
30. White, I.R., Royston, P., Wood, A.M.: Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine **30**(4), 377–399 (2011)
31. Zapata, M.E., Rovirosa, A., Carmuega, E.: Cambios en el patrón de consumo de alimentos y bebidas en argentina, 1996-2013. Salud colectiva **12**, 473–486 (2016)