

Métodos de recolección y precisión de encuestas. El caso de Nuevo León, México en 2021

Mode Effects and Survey Accuracy. The Case of Nuevo León, Mexico in 2021

Alejandro Díaz Domínguez*

Reibido el 20 de abril de 2023
Aceptado el 27 de junio de 2023

Resumen

Al analizar los diferentes métodos de recolección de datos por encuesta para la elección a la gubernatura de Nuevo León en 2021, se encontró, en consonancia con la literatura, que el error promedio se redujo conforme se acercaba la fecha de la elección, si se consideraba un horizonte temporal amplio. En algunos casos, las características metodológicas de las encuestas, como el margen de error, el tamaño de muestra y el número de días en campo, también ayudaron a reducir dicho error. Sin embargo, según el método de recolección empleado, en el que los métodos probabilísticos resultaron relativamente más acertados que los no probabilísticos, el error promedio y el error de estimación por candidatura presentaron variaciones en su precisión.

Palabras clave

Encuestas, elecciones estatales, opinión pública, México, Mosteller.

* Doctor en Ciencia Política por la Universidad de Vanderbilt. Actualmente es profesor investigador de la Escuela de Gobierno del Tecnológico de Monterrey. alejandrod.dominguez@tec.mx. ORCID: <https://orcid.org/0000-0002-3856-5686>

Abstract

When analyzing survey data collected by different methods for the 2021 gubernatorial elections in Nuevo León, Mexico, it was found, in line with the literature, that the average error decreased as the election day approached, when considering a broad time horizon. In some cases, polls' methodological characteristics also reduced the error, such as margin of error, sample size and the number of days spent in the field. However, according to the collection method used, in which probabilistic methods were relatively more accurate than the non-probabilistic ones, the average error and the estimation error per candidate showed variations in accuracy.

Keywords

Polls, state elections, public opinion, Mexico, Mosteller

¿Qué tanto se acercan las estimaciones de encuestas prelectorales a los resultados finales según el método de recolección? ¿Qué tanto influye el tiempo previo en la exactitud de las encuestas respecto del resultado final? Para intentar responder a estas preguntas se estudia la literatura sobre exactitud de encuestas, así como se verifica la influencia de los distintos métodos de recolección y el tiempo previo a la jornada electoral en la exactitud de las encuestas mediante regresiones lineales. Para ello, se compila una base de datos que contiene 155 mediciones públicas de 17 casas encuestadoras para las elecciones a la gubernatura de Nuevo León en 2021.

Los análisis empíricos sobre el impacto que tienen los diversos métodos de recolección en la exactitud de las encuestas preelectorales son aún escasos en literatura y más en el ámbito subnacional. De ahí que resulte pertinente estudiar si los diversos métodos inciden de manera diferenciada en dicha exactitud y si conforme se acerca la jornada electoral, los estudios demoscópicos se acercan en sus estimaciones al resultado final. En este ejercicio se halla que sí existen diferencias en los métodos de recolección, puesto que los probabilísticos resultan más acertados, mientras que, con independencia del método, la cercanía con la fecha de la elección sí se asocia positivamente con una mayor exactitud en las estimaciones por encuesta.

Es posible observar diferencias en las estimaciones que tienen distintos métodos para recopilar datos por encuesta, debido a la diversidad de públicos que cubren. Otras fuentes de variación son el diseño de la muestra, el orden de las preguntas en el cuestionario, la duración de la entrevista, los días en los cuales se lleva a cabo el levantamiento (entre semana, fines de semana o ambos), el número de entrevistas y el margen de error. Si examinamos distintos métodos de recolección de datos en encuestas durante una misma elección, podemos observar diferencias de precisión en sus resultados. Las diferencias metodológicas en técnicas de muestreo, número de días de trabajo en campo y estrategias de reclutamiento podrían ayudar a entender mejor dichas variaciones. Además, una virtud del ejercicio demoscópico en elecciones es que está sujeto a la evaluación del dato poblacional, el cual se compara con el resultado electoral final, una característica que no se presenta con frecuencia en la investigación social.

En la elección a la gubernatura de Nuevo León en 2021 fue una de las primeras ocasiones donde se observó una mayor diversidad en métodos de recolección, pues se aplicaron cinco modalidades diferentes. Adicionalmente, en ella se presentó un vuelco en las preferencias, ya que el candidato que se ubicaba en el cuarto lugar al inicio de la contienda terminó ganando las elecciones (de la Garza, 2021). Si bien es cierto que no es la primera vez que ocurre un vuelco así, pues ya se había visto en Nuevo León en 2015 y en el Estado de México en 2017 (Moreno, 2018, pp. 156-158), dicho cambio en las preferencias, sumado a la diversidad de métodos empleados, convierte a la elección de Nuevo León en 2021 en un caso relevante para el estudio de los métodos de recolección de datos.

Revisión de la literatura

Las variaciones en los resultados de mediciones por encuesta han sido objeto de diversos análisis en el caso mexicano. Cuando el foco se ha centrado en las elecciones nacionales, se han analizado el impacto de los efectos de casa y la incertidumbre existente en las estimaciones a través de filtros de Kalman (Valle-Jones, 2012; Cantú et al., 2016), o bien, los efectos de casa en mediciones preelectorales en el contexto de la campaña presidencial (esto es, después de cada debate), así como la precisión de las encuestas de salida (De la Peña, 2020).

Cuando se ha analizado la precisión de las encuestas preelectorales a nivel estatal, se ha incluido el orden de la pregunta de intención de voto, si hubo sustitución de puntos de muestra, el énfasis en muestras urbanas *versus* rurales, el espiral del silencio, el contexto de inseguridad, el género, la edad y la experiencia de quien aplica el cuestionario, la supervisión y la duración de la entrevista en una encuesta preelectoral estatal en vivienda con diversos tratamientos. Se halló que las submuestras con mayor cobertura rural, realizadas por entrevistadoras y con supervisión, redujeron el error de estimación; mientras que, donde se percibió alguna presión durante la entrevista, se apreció un error mayor (Moreno et al., 2014).

En compilaciones de encuestas estatales, se ha encontrado que importan más los efectos de casa que las no respuestas y que diversos elementos del contexto son significativos, como la incidencia delictiva o la competencia partidista (Romero, 2012). Desde entonces ya se delineaban como posibles fuentes de error de medición el orden de la pregunta de intención de voto, si se empleó boleta (o si fue pregunta directa), el tipo de muestreo y la calidad en el trabajo de campo (Romero, 2012, pp. 118-121).

En los diferentes esfuerzos de análisis pareciera necesario incluir el método de recolección, ya que la vasta mayoría de los estudios analizados para el caso mexicano se han centrado en levantamientos en vivienda y en telefónicas en vivo (Moreno, 2018, pp. 153-154) y, en menor medida, en encuestas probabilísticas por la red social Facebook y telefónicas por robot (De la Peña, 2020). Adicionalmente, la proliferación de encuestas no probabilísticas con la pandemia por COVID-19, pudiera dar pie a entender mejor el papel que juegan los diversos métodos de recolección en encuestas preelectorales.¹

La literatura reporta la existencia de algunas diferencias consistentes según el método de recolección, ya que se ha encontrado que las encuestas en línea tienen menos probabilidades de detectar movimientos a lo largo de la campaña electoral que las encuestas telefónicas o las telefónicas por robot, también llamadas

1 La pandemia ocasionada por el SARS-CoV-2 limitó las posibilidades de levantar encuestas en vivienda, debido a las reglas de distanciamiento en las interacciones presenciales. El número de encuestas levantadas en vivienda entre marzo y diciembre de 2020 fue de solo 24 encuestas nacionales y principalmente a partir de agosto, con apenas 2 encuestas estatales, una a finales de agosto y otra a mediados de octubre en Nuevo León y Sonora. De las 38 encuestas nacionales entre marzo y diciembre de 2020 incluidas en el agregado de aprobación presidencial de *Oraculus*, 25 fueron telefónicas (Díaz Domínguez, en prensa).

IVR - *Interactive Voice Response*. En ocasiones, los levantamientos IVR pueden tener mayor precisión, pero tienden a sobreestimar a las opciones políticas conservadoras (Durand y Johnson, 2021).

Los estudios para el caso mexicano han encontrado mayores variaciones en el ámbito local, pero se han centrado en uno o dos métodos de recolección. Escasos análisis han incorporado factores metodológicos más allá del efecto de casa, por lo que puede resultar de utilidad verificar el papel que han jugado los distintos métodos de recolección a lo largo de una campaña electoral estatal. Adicionalmente, debe recordarse que las encuestas a nivel entidad federativa para elecciones a gubernaturas han registrado históricamente, en promedio, un mayor error de estimación que las federales (Romero, 2012; Moreno, 2018, p. 232).

Dentro de los factores metodológicos que pueden ayudar a entender variaciones en la precisión de los ejercicios de encuesta destacan el número de días antes de la elección, el tamaño de muestra y los métodos de recolección. El tamaño de muestra se ha relacionado con sesgos en la medición (Panagopoulos et al., 2018, p. 167), porque un mayor tamaño de muestra y un menor margen de error incrementan la exactitud (Sohlberg, 2017, p. 5; Rich et al., 2018, pp. 131-135). Respecto al horizonte temporal, la inclusión de mediciones más alejadas supone que sea posible verificar si el efecto del tiempo se traduce en mejores estimaciones conforme se acerca la jornada electoral (Jennings y Wlezien, 2018; Panagopoulos et al., 2018, p. 166).

Si bien el análisis sobre la exactitud de las encuestas generalmente se basa en la última semana antes de las elecciones (Moreno, 2018), no es inusual encontrar evaluaciones con horizontes temporales mayores. Por ejemplo, la *American Association of Public Opinion Research* (AAPOR) analizó el papel de las encuestas en 2016, considerando las últimas dos semanas (Kennedy, 2017). De ahí que se incluyan encuestas más lejanas, esto es, a una, dos, cinco o hasta seis semanas antes del día de las elecciones (Romero, 2012; Cantú et al., 2016; Jennings y Wlezien, 2018; Sohlberg et al., 2017; Panagopoulos et al., 2018).

En vista de que un horizonte temporal mayor ayuda a distinguir los errores por aleatoriedad de los errores sistemáticos (Cantú et al., 2016), en el presente estudio se propone abarcar un margen de cinco meses y medio. Ello permitirá verificar si los diferentes métodos de recolección también reducen el error conforme se aproximan las elecciones.

Durante la contingencia sanitaria, que continuó durante las campañas electorales de 2021, las muestras no probabilísticas ganaron espacio debido a sus múltiples ventajas: no solo resultan relativamente más baratas, sino que también son de rápido procesamiento y cuentan con una base de personas a entrevistar previamente integrada por panelistas voluntarios, la cual puede ser filtrada para evitar la inclusión de seguidores interesados (Castorena et al., 2023, p. 279). En los últimos años, en las elecciones mexicanas se han empleado diversos métodos de recolección, como las encuestas que seleccionan personas entrevistadas a través de redes sociales como Facebook; las levantadas vía telefónica (celular o residen-

cial), con entrevista en vivo o por robot (IVR); las efectuadas a través de mensajería instantánea (como WhatsApp, que bien podrían ser catalogadas como *opt-in*) y, finalmente, las que realizan la entrevista en vivienda, cara a cara.

Las encuestas de corte probabilístico son instrumentos que miden, con base en una muestra representativa, características de la población, como podrían ser las preferencias políticas. Para que las encuestas permitan inferir que los hallazgos derivados de la muestra son aplicables a la población, estas necesitan ser probabilísticas, esto es, se requiere que las personas entrevistadas sean seleccionadas al azar (Rich et al., 2018). Dentro de este grupo pueden considerarse los métodos en vivienda y los levantamientos telefónicos en vivo.

La recolección de datos en vivienda se realiza a través de la visita de una persona que aplica el cuestionario en casa. Se confecciona una muestra con las secciones electorales, se eligen puntos de muestra al azar y se aplica una decena o docena de cuestionarios en las viviendas aledañas. Es importante que exista cierta dispersión, es decir, que no se realicen entrevistas en casas muy juntas; para ello se emplean diversas técnicas, como la del caracol (Rich et al., 2018, pp. 122-130). Los datos pueden levantarse de forma tradicional, esto es, en papel, o por medio de un dispositivo electrónico, como un PDA (*Personal Digital Assistant*), un teléfono inteligente o una tableta. Dos retos que enfrentan los encuestadores que siguen este método son 1) el sesgo de deseabilidad social, es decir, cuando una persona responde en una encuesta algo por considerarlo socialmente aceptable, aunque no esté de acuerdo; y 2) la falta de respuesta (Durand y Johnson, 2021, p. 184), la cual da pie a problemas conocidos en la literatura como la espiral del silencio.

Las encuestas telefónicas en vivo se levantan a través de llamadas a teléfono fijo y celular, y su nivel de penetración es dos veces mayor que el residencial (Moreno, 2018). Por lo general, se confecciona un universo de teléfonos a través de la generación de números al azar. Después, se elabora la muestra según las claves de llamada, las cuales están vinculadas a diversas zonas del país. Durante la llamada, un entrevistador invita a la persona entrevistada a contestar una serie de preguntas (Rich et al., 2018, pp. 150-151). En esta modalidad los cuestionarios son un poco más cortos, en comparación con los levantados en vivienda. Y, al igual que estas últimas, también se enfrentan al sesgo de deseabilidad social.

Las encuestas telefónicas levantadas a través de un robot formulan preguntas con una voz grabada, invitan a la persona encuestada a teclear los números seleccionados en su teléfono digital y registran las respuestas proporcionadas. Los cuestionarios son mucho más cortos en comparación con los levantados por las encuestas telefónicas en vivo (De la Peña, 2020; Díaz Domínguez, en prensa). Y, si bien estas resultan más baratas, es común que las encuestas telefónicas realizadas en vivo no sean completadas en su totalidad, por lo que suelen requerir mayores esfuerzos por parte del personal, esfuerzos que, en el caso de las encuestas levantadas por un robot, simplemente se transfieren a un *software* (Rich et al., 2018, pp. 156-157).

Las encuestas telefónicas con robot tienen algunas limitaciones. Por ejemplo, la reducida extensión del cuestionario no ofrece un perfil demográfico detallado de la muestra en estudio y tampoco permite conocer opiniones en torno a temas diversos, dentro de una misma muestra; lo cual no ocurre con las encuestas en vivienda o las realizadas en línea (correo electrónico, Facebook o WhatsApp). Lo anterior impide una posible postestratificación, ya que no permite realizar algunos cruces, debido a la limitada variedad de temas (De la Peña, 2020, p. 126). Sin embargo, este sería un aspecto que se puede perfeccionar a medida en que se empleen cuestionarios más extensos para temas específicos, mientras no excedan la decena de preguntas. Finalmente, cuando las telefónicas de robot comienzan a distanciarse del conjunto de encuestas, en ocasiones recurren a ajustes inesperados que las acercan al promedio de otros ejercicios. Este fenómeno se conoce como *herding* (Durand y Johnson, 2021, p. 185).

Por otro lado, los mayores niveles de educación e ingreso que se presentan en las encuestas telefónicas residenciales, así como las que emplean correo electrónico o Facebook, pudieran tener un impacto diferenciado en variables como la aprobación de desempeño de gobierno, intención de voto o conocimiento de figuras políticas. Otro componente que varía es la edad, ya que el alcance pudiera ser mayor: en adultos jóvenes, cuando se trata de telefonía celular y Facebook, y en personas mayores, cuando se hacen levantamientos mediante teléfono residencial, pues dicho segmento hace menor uso de la mensajería instantánea o IVR (Durand y Johnson, 2021, p. 186).

Ante las diferentes coberturas, en el caso de las telefónicas, las casas encuestadoras realizan diversas combinaciones entre encuestas residenciales y a celular, con miras a incluir otros segmentos de la población que solo con teléfono residencial no se alcanzarían. Un reto en todo método de recolección es reducir la tasa de rechazo. Las telefónicas en vivo han reportado una tasa de rechazo del 40 y hasta el 55 por ciento, mientras que las de vivienda muestran rechazos de entre 38 y 44 por ciento. Las *opt-in* lo reportan hasta el 80 por ciento, mientras que las telefónicas por robot no suelen reportar este dato (Díaz Domínguez, en prensa).

Las encuestas levantadas en línea (*web surveys* u *opt-in*) se optimizan para ser respondidas en computadoras, tabletas y teléfonos inteligentes, para que las personas puedan contestar el cuestionario accediendo a algún servicio de internet. Esta modalidad se construye a partir de paneles de personas voluntarias que, a la larga, pueden convertirse en personas casi profesionales en responder encuestas (Rich et al., 2018, pp. 152-153). Por esta razón, teóricamente, este tipo de encuestas no deberían reportar margen de error, ya que carecen de un marco muestral, aunque sí es posible conocer la tasa de rechazo. Si bien, no se ha encontrado un problema mayor en torno al sesgo de deseabilidad social, (Antoun et al., 2017, pp. 283-284, 295), la falta de optimización en teléfonos inteligentes puede inducir a errores (Antoun et al., 2017, p. 299).

Su funcionamiento general pudiera describirse de la siguiente manera: se contacta a las personas por redes sociales o correo electrónico para configurar

un número suficiente de entrevistas, el cual se pondera con factores demográficos como estrategia de postestratificación, con el objetivo de que se asemeje a las proporciones de la población en estudio, según datos de reportes oficiales o censales. A este conjunto de pasos se le conoce como ponderación de calibración y sirve para minimizar sesgos por no respuesta, aunque también permite lidiar con posibles sesgos derivados de no incluir a proporciones relevantes de la población en estudio. La calibración puede realizarse en uno o dos pasos para minimizar la variabilidad (Kott y Liao, 2015), o bien, puede hacerse de manera mixta, esto es, mezclando una muestra probabilística ponderada con una no probabilística (Cornesse et al., 2020). Los ejercicios de postestratificación mejoran muy poco la calidad de las muestras (Castorena et al., 2023, pp. 274-277), pues sobrerrepresentan a la población con mayores niveles de ingreso y escolaridad, que son también quienes tienen un mayor acceso a internet.

La recolección de datos a través de la red social Facebook opera a través de un universo limitado a personas usuarias de dicha red social. Delimita un perfil, por ejemplo, 18 años y más, establece determinadas proporciones por género, grupos de edad y nivel de escolaridad. Con esa base, diseña una muestra o trata los datos posteriormente con métodos de calibración. Una vez definida la lista, se envía un cuestionario a cada persona seleccionada; al lado de su muro aparecerá un anuncio que invita al usuario a contestar la encuesta. El ejercicio termina cuando se cubren los números absolutos requeridos, de conformidad con las proporciones del perfil previamente delimitado.

Las encuestas por mensajería instantánea forman parte de los métodos no probabilísticos, ya que carecen de marco muestral. Por tal razón, la lista de personas a entrevistar se construye con la compilación de numerosas bases de datos que permite enviar encuestas a la unidad territorial deseada a través de la clave telefónica estatal correspondiente, la cual se ajusta según factores demográficos. Ante la falta de una muestra, no existe una selección azarosa de las personas a entrevistar. Sin embargo, a pesar de esta construcción no aleatoria, el margen de error que usualmente se reporta parece estimarse asumiendo supuestos teóricos de muestreo en los cuales la selección sí se efectúa al azar, a diferencia de las encuestas por WhatsApp.² En este método se depende de celulares inteligentes con conexión a internet en algún momento del día para responder la encuesta (Díaz Domínguez, en prensa).

La literatura reporta que las encuestas en línea, también llamadas *opt-in* pueden resultar menos exactas que las probabilísticas, porque requieren de ajustes para asemejar sus estimaciones a la distribución poblacional (Sohlberg et al., 2017,

2 Generalmente se supone una varianza conocida como máxima en una población infinita (la conocida igualdad $q=1-p$, con $p=0.5$, lo cual tomaría un intervalo muy amplio con un tamaño de muestra enorme, es decir, infinito). Este tipo de supuestos sirven como comparación cuando se cuentan con estimaciones anteriores recientes o muestras piloto para contar con una aproximación del valor p . Lo que parece ocurrir en estos casos es que, al tratarse de muestras no probabilísticas, el adoptar supuestos para muestras que provienen de selecciones al azar pudiera resultar controversial si no se explica detenidamente la construcción del margen de error (Saris y Gallohofer, 2014; Rich et al., 2018).

p. 6). Dicho ajuste puede realizarse a través de diversas técnicas, como la calibración, el *raking* o la ponderación *ad hoc* (Hartman y Levin, 2019). Las encuestas no probabilísticas han resultado menos exactas, como lo sugieren algunos análisis de *opt-in*, en comparación con encuestas levantadas en vivienda (Sohlberg et al., 2017; Cornesse et al., 2020).

Por un lado, las encuestas en línea no alcanzan a cubrir a la población que carece de acceso a internet, la cual generalmente es de mayor edad, reporta menores niveles de ingresos y escolaridad, y vive en localidades rurales (Sohlberg et al., 2017, pp. 3-4; Castorena et al., 2023, p. 274). En consecuencia, este perfil puede diferir del que sí cuenta con acceso a la red. Por otro lado, el reclutamiento es voluntario, lo que también puede inclinar la muestra hacia personas con mayor interés en la política, mayor consumo de noticias y participación electoral, características que no necesariamente comparten con la población abierta (Sohlberg et al., 2017, p. 4).

Desde luego, el problema de una amplia tasa de rechazo afecta a cualquier método de recolección, pero la variación en dicha tasa parece relacionarse con la forma en la cual se recopilan los datos. En otras palabras, las encuestas en línea o por mensajería instantánea, así como las telefónicas levantadas a través de robot, parecen tener tasas más altas que las telefónicas en vivo y las recopiladas en vivienda. Lo anterior implica que se requieren de numerosos intentos adicionales para alcanzar un número razonable de entrevistas (Sohlberg et al., 2017, p. 5).³

Finalmente, para evaluar la exactitud de las encuestas, se emplean métricas como el error promedio, calculado con el valor absoluto del error de cada candidatura entre el total de candidaturas. Esta métrica es conocida como M3 o Mosteller 3 (Mosteller et al., 1949). Otros estudios emplean A' , una medición que incorpora los sesgos sobre un partido en relación con otro (Martin et al., 2005), el cual se esperaría con valor cero (Panagopoulos et al., 2018, pp. 163-164), o el indicador compuesto B (Arzheimer y Evans, 2014), el cual deriva de un modelo multinomial ponderando por el tamaño de cada partido, lo cual lo convierte en una métrica atractiva para sistemas multipartidistas. Esta última ponderación intenta aliviar el problema de M3, el cual asume que todas las candidaturas valen lo mismo cuando se divide la suma de los errores absolutos entre el total de candidaturas. Sin embargo, algunos estudios sugieren que no parecen apreciarse diferencias sustantivas entre M3 y B (Sohlberg et al., 2017, p. 10).

Por las razones antes mencionadas, este artículo se basará en M3 y en el error por cada candidatura, considerando la diferencia que se extrae de restar el porcentaje estimado por cada medición de encuesta y el dato oficial, con objeto de conocer las sub y sobreestimaciones, en línea con lo propuesto por M6 (Fien-

3 Si bien es cierto que los cálculos para el tamaño de muestra, el nivel de confianza y el margen de error se pueden basar en supuestos del teorema central del límite para el caso de muestras probabilísticas, también es cierto que otros factores de diseño afectan el tamaño de muestra, como los niveles de homogeneidad y el número de categorías a investigar. Ante limitaciones en tiempo y recursos, modos de recolección diferentes a levantamientos en vivienda requieren que la discusión estadística y implicaciones prácticas se efectúe al mismo tiempo, ya que emplean muestras no probabilísticas (Rich et al., 2018).

berg y Hoaglin, 2006). Un problema con esta aproximación es que deja de lado el agregado, privilegiando una sola medición (De la Peña, 2020). Ante el dilema que resulta de las objeciones a cada método mencionado y la siempre bienvenida simplicidad del método de evaluación, se emplearán M3 y M6.

Hipótesis, datos y métodos

Con base en la revisión de la literatura sobre métodos de recolección, se formulan las dos hipótesis siguientes:

H1: Los métodos de recolección con entrevista en vivo (*interviewer - assisted*), como las encuestas en vivienda y las telefónicas en vivo, presentarán un menor error promedio y menores errores de estimación por candidatura.

H2: Los métodos de recolección no probabilísticos (IVR y *opt-in*), como telefónicas por robot, mensajería instantánea a través de WhatsApp y la red social Facebook, presentarán un mayor error promedio y mayores errores de estimación por candidatura.

Asimismo, se espera que todos los métodos de recolección, conforme se acerca la fecha de la jornada electoral, reduzcan sustancialmente su error de estimación, esto es, que el porcentaje de intención de voto estimado, en relación con el resultado oficial, se ubique dentro de sus propios márgenes de error (Moreno, 2018, p. 202). De igual forma, se espera que un menor margen de error se asocie con un menor error promedio y con menores errores de estimación por candidatura. Respecto al tamaño de muestra, que para este ejercicio sea estandarizado en miles, se espera que un mayor número de casos también reduzca errores de estimación, tanto promedio como por candidatura. Finalmente, respecto al número de días en campo, se asumirá que más días de levantamiento pueden reducir ambos tipos de errores.⁴

Un ejercicio aplicado a las elecciones a la gubernatura en Nuevo León podría ilustrar las diferencias entre los distintos métodos de recolección, dado que se trató de una elección subnacional donde se observaron cinco métodos de recolección. Se analizaron 155 mediciones entre el 17 de diciembre de 2020 y el 1 de junio de 2021 para cuatro candidaturas a la gubernatura de Nuevo León: Clara Luz Flores, quien obtuvo el 14.0264 por ciento de los votos y fue postulada por MORENA, Partido Verde Ecologista de México, Partido del Trabajo y Partido Nueva Alianza Nuevo León; Adrián de la Garza con 27.9071 por ciento, postulado por los

⁴ Debe matizarse que la suposición sobre la cercanía de la jornada electoral como variable que reduce tanto el error promedio como por candidatura, se refiere al resultado final, sin prejuzgar sobre los vaivenes o efectos de campaña, sobre los cuales no se cuenta con el parámetro poblacional.

partidos Revolucionario Institucional y de la Revolución Democrática; Fernando Larrazábal con 18.3341 por ciento, lanzado por el Partido Acción Nacional, y Samuel García con 36.7151 por ciento, al frente de Movimiento Ciudadano.⁵

Las mediciones compiladas fueron levantadas por las casas Arias Consultores, Buendía & Márquez, Campaigns & Elections, De las Heras, Demoscopia Digital, Moreno & Sotnikova Social Research and Consulting para *El Financiero*, *El Norte*, Enkoll, Factométrica, *El Heraldo*, *El Horizonte*, Massive Caller, México Elige, Poligrama, *Reforma*, Varela y Asociados y *Voz y Voto*. Los métodos de recolección analizados fueron: Facebook con 13 mediciones, telefónica por robot con 101, telefónica en vivo con 11, vivienda con 15 y WhatsApp con 15. Los datos provienen del repositorio institucional de encuestas de la Comisión Estatal Electoral (CEE, 2021a) y de los sitios de internet y redes sociales de las casas encuestadoras citadas.⁶ Finalmente, los resultados oficiales de la elección a la gubernatura provienen de los cómputos posteriores a las impugnaciones (CEE, 2021b).

Como variables dependientes se emplearon dos tipos de errores: M3 y M6. Para el caso de M3, esto es, el error promedio en valores absolutos, se calcularon las diferencias en valores absolutos entre el porcentaje estimado y el dato oficial para cada una de las cuatro candidaturas, dichos valores se sumaron y se dividieron entre cuatro. Para el caso de M6, el error de estimación por candidatura, simplemente se calcularon las diferencias entre el porcentaje estimado y el dato oficial para cada una de las cuatro candidaturas, donde los valores positivos indican sobreestimación y los valores negativos indican subestimación. Cuando M3 y M6 tienen un valor muy cercano al cero, se trata de una mejor estimación.

Las variables explicativas son el margen de error, el tamaño de muestra dividido entre mil, el número de días en campo, el número de días antes de la jornada electoral y variables binarias que identifican a cada método de recolección. Dado que, exceptuando los de robot, los métodos de recolección restantes solo presentan alrededor de 50 mediciones, no fue posible incluir los efectos de casa como efectos fijos, aunque se asumirá que las características metodológicas pudieran servir como controles relativamente eficaces ante la falta de los efectos mencionados. La estadística descriptiva se muestra en la Tabla 1.⁷

Se estimaron dos modelos lineales para M3, uno como base y otro interactivo, donde se combinó cada método de recolección con los días restantes para la

5 La contienda fue entre siete candidaturas, las cuatro mencionadas más Carolina Garza, con 0.3286 por ciento (PES), Daney Siller con 0.3093 por ciento (RSP) y Emilio Jacques, con 0.6468 por ciento (FXM). Numerosas estimaciones les calculaban uno por ciento o menos, de ahí que por simplicidad no se incluyeron en la base de datos ni en las métricas de evaluación.

6 La encuesta levantada por Moreno & Sotnikova el 25 de abril de 2021 fue reportada como mixta, al contar con un componente en vivienda (800 entrevistas) y otro con teléfono en vivo (400 entrevistas). Al ser la única mixta en términos de métodos de recolección, en la base de datos se consideró como telefónica en vivo, de acuerdo con el método de recolección de la referida casa encuestadora. Si se cataloga como en vivienda, los resultados del presente análisis no varían sustancialmente.

7 La tasa de rechazo no pudo ser incluida como variable explicativa porque en 114 casos no se reportó dicho dato, principalmente por parte de las encuestas telefónicas levantadas mediante robot.

jornada electoral, de modo que se pudiera observar si se captaban movimientos derivados de la dinámica de las campañas y si, en efecto, el error promedio disminuía al acercarse el día de los comicios. Para el caso de M6 también se emplearon modelos lineales, se estimaron cuatro modelos base, uno por candidatura y cuatro modelos interactivos; también uno para cada candidatura, donde se combinó cada método de recolección con los días para la jornada electoral. Los modelos interactivos con un determinado horizonte temporal han resultado una estrategia efectiva para analizar, tanto la precisión, como los movimientos de campaña captados por cada método de recolección (Durand y Johnson, 2021).

Para mostrar los vaivenes de la campaña electoral, la Gráfica A1 (disponible en el anexo en línea) muestra 155 mediciones a lo largo de los cinco meses y medio, con ajustes de regresiones locales para cada una de las cuatro candidaturas, con el dato oficial en el día cero, donde cada círculo representa una medición de encuesta. Al principio de la campaña, la candidata Clara Luz Flores (MORENA-PT-PVEM-NA) encabezaba las preferencias, seguida por el candidato Adrián de la Garza (PRI-PRD), mientras que en tercer lugar se observaba al candidato Fernando Larrazábal (PAN) y, en un lejano cuarto lugar, al candidato Samuel García (MC). Hasta entonces el panorama lucía claro.

Tabla 1. Estadística descriptiva

Variable	Promedio	Desv. est.	Mínimo	Máximo
Error promedio (M3)	0.07	0.03	0.01	0.15
Error de la Garza	-0.001	0.04	-0.10	0.12
Error García	-0.10	0.10	-0.28	0.07
Error Larrazábal	-0.03	0.04	-0.10	0.08
Error Flores	0.08	0.06	-0.04	0.24
Muestra (en miles)	1.09	0.63	0.40	6.70
Margen de error	0.03	0.01	0.01	0.05
Días en campo	2.17	3.27	1	28
Días para la jornada	63	42	5	171
Método Facebook	0.08	0.28	0	1
Método telefónico robot	0.65	0.48	0	1
Método telefónico en vivo	0.07	0.26	0	1
Método vivienda	0.10	0.30	0	1
Método WhatsApp	0.10	0.30	0	1

Fuente: Compilación propia. 155 mediciones de 17 casas encuestadoras entre el 17 de diciembre de 2020 y el 1 de junio de 2021 (CEE, 2021a), redes sociales y páginas de internet de las propias casas encuestadoras. Los resultados oficiales fueron obtenidos después de las impugnaciones (CEE, 2021b), aunque no variaron con respecto a los cómputos distritales, ya que dichas impugnaciones apuntaban al financiamiento y no al proceso de votación.

Sin embargo, conforme avanzó la campaña, los escándalos no se hicieron esperar y el panorama que resultaba claro dejó de serlo. Primero, la candidata Clara Luz Flores presentó demandas contra un *youtuber* debido a que se sintió agraviada por uno de sus videos; y, a los pocos días, se dio a conocer una grabación (difundida por el PRI) donde la propia candidata aparecía con el líder del censurable grupo NXIVM durante la tercera semana de marzo (de la Garza, 2021, p. 144). Si bien circulaban en redes sociales numerosas fotografías y testimonios de diversos integrantes de la clase política mexicana que habían acudido a cursos de la organización citada, en algunas entrevistas, incluso ese mismo día, la candidata había negado cualquier vínculo con dicho grupo. El golpe mediático pareció tener efectos: en la Gráfica A1, se aprecia que la candidata de la coalición encabezada por MORENA comenzó a descender en las preferencias para terminar en cuarto lugar.⁸

Lo que ocurrió después fue objeto de debate durante algunas semanas: se quería determinar si el líder de la carrera a la gubernatura era el candidato de la coalición encabezada por el PRI o el candidato panista, ello para que se diera la declinación de uno en favor del otro. También se debatía si el candidato de MC habría repuntado lo suficiente como para empezar a disputar el primer lugar. Aunque al final la distancia entre García (36.7151) y de la Garza (27.9071) fue de 8.808 puntos, diversos medios auguraban una contienda mucho más cerrada (menos de dos puntos) al emplear varias de las mediciones que también se analizan en el presente estudio (Galindo, 2021).

Todos estos vaivenes sugieren que las elecciones a la gubernatura en Nuevo León representan una buena oportunidad para analizar el desempeño de las encuestas y de los métodos de recolección en un horizonte temporal amplio. Además, permiten observar si entre más cercana es la fecha de las elecciones, el error de estimación se reduce por tipo de método de recolección y si dichos métodos captan los movimientos derivados de los vaivenes electorales ya descritos.

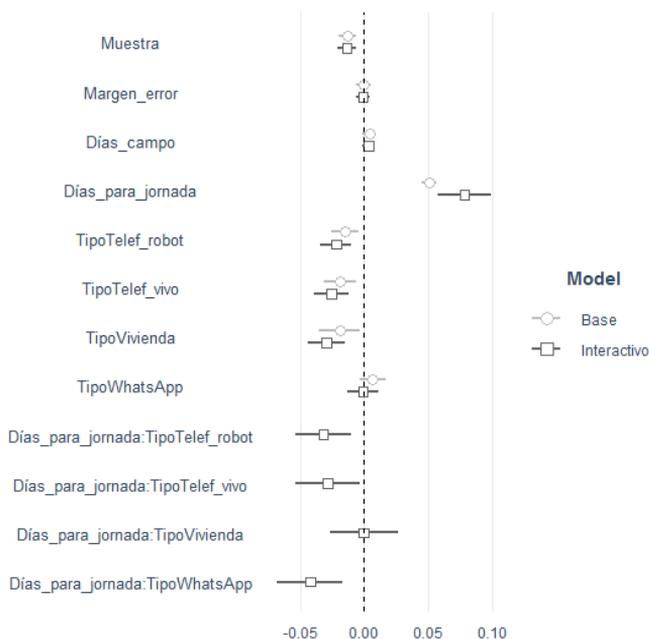
La Gráfica A2 (disponible en el anexo en línea) muestra el error promedio en dos horizontes temporales: cinco meses y medio y dos semanas previas a la elección. En ambos se aprecia que, en efecto, el error se reduce conforme se acerca el día de los comicios y, en particular, dentro del horizonte de dos semanas se observa que el error promedio ya es inferior a 5 puntos. Todas estas piezas preliminares de información sugieren que los presupuestos iniciales sobre la temporalidad se verifican. A continuación, se examinará el papel que jugaron los métodos de recolección.

8 Quedará para investigaciones posteriores explorar si el ser mujer influyó en que la candidata fuera juzgada de esa forma y que haya experimentado tal caída; ya que, en sociedades relativamente conservadoras, este tipo de situaciones pudieran ser juzgadas con un rasero diferente según el género.

Resultados

Para conocer qué variables se asocian con el error promedio, la Gráfica 1 muestra dos modelos lineales: uno que sirve de base y otro que es interactivo, donde se combinan los días para la jornada electoral y cada uno de los métodos de recolección. En el modelo que sirve de base, se observa que muestras con más entrevistas, la cercanía con la jornada comicial, los métodos telefónicos (en vivo y por robot) y las encuestas levantadas en vivienda ayudan a reducir el error promedio. Una muestra de 400 casos tendría un error promedio de 9 y una de mil de 7.5 puntos. Respecto al horizonte temporal, en diciembre el error era de 15 puntos, según este modelo, tres meses antes era de 9, dos meses antes de los comicios de 7.5, un mes antes de 6, 15 días antes de 5 y cinco días antes de 4.8.

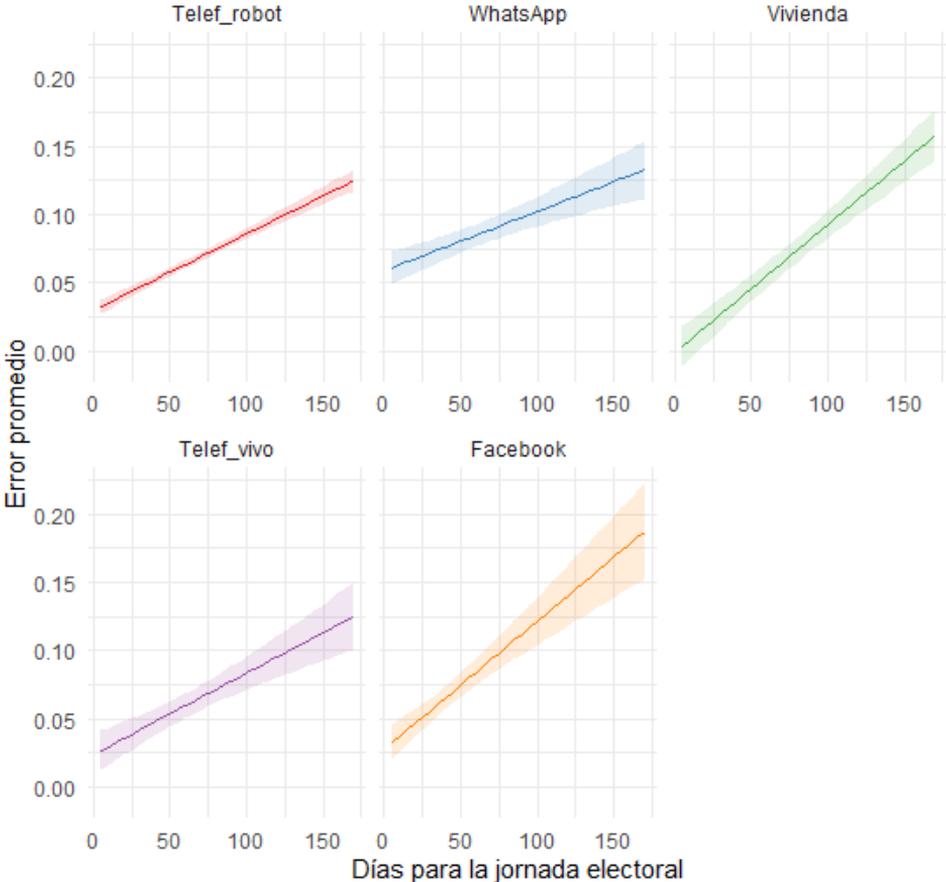
Gráfica 1. Determinantes del error promedio (M3)



Fuente: Elaboración propia. Estimaciones lineales. Variable dependiente: error promedio (M3). El modelo base se identifica con círculos y el interactivo con cuadrados, el cual agrega cuatro combinaciones entre cuatro métodos o tipos de recolección y los días hacia la jornada electoral. El método de recolección por Facebook sirve como categoría de referencia. Modelo base: 155 casos, R2: 0.66, prueba F: 38.93, error de la regresión 0.02. Modelo interactivo: 155 casos, R2: 0.69, prueba F: 30.09, error de la regresión: 0.02. Las líneas identifican errores estándar corregidos al 90 por ciento y las figuras identifican a los coeficientes. Aquellos que tocan la línea vertical cero no son significativos, aquellos ubicados a la izquierda se asocian negativamente y aquellos ubicados a la derecha de la línea cero se asocian positivamente.

Finalmente, los métodos de recolección, vivienda y telefónica en vivo presentan un error promedio de 6 puntos, las telefónicas por robot de 6.5, el levantamiento por Facebook de 8 y las encuestas por WhatsApp de 8.6, como se aprecia en la Gráfica 2.

Gráfica 2. Predicción de probabilidades de error promedio, según días para la jornada y métodos



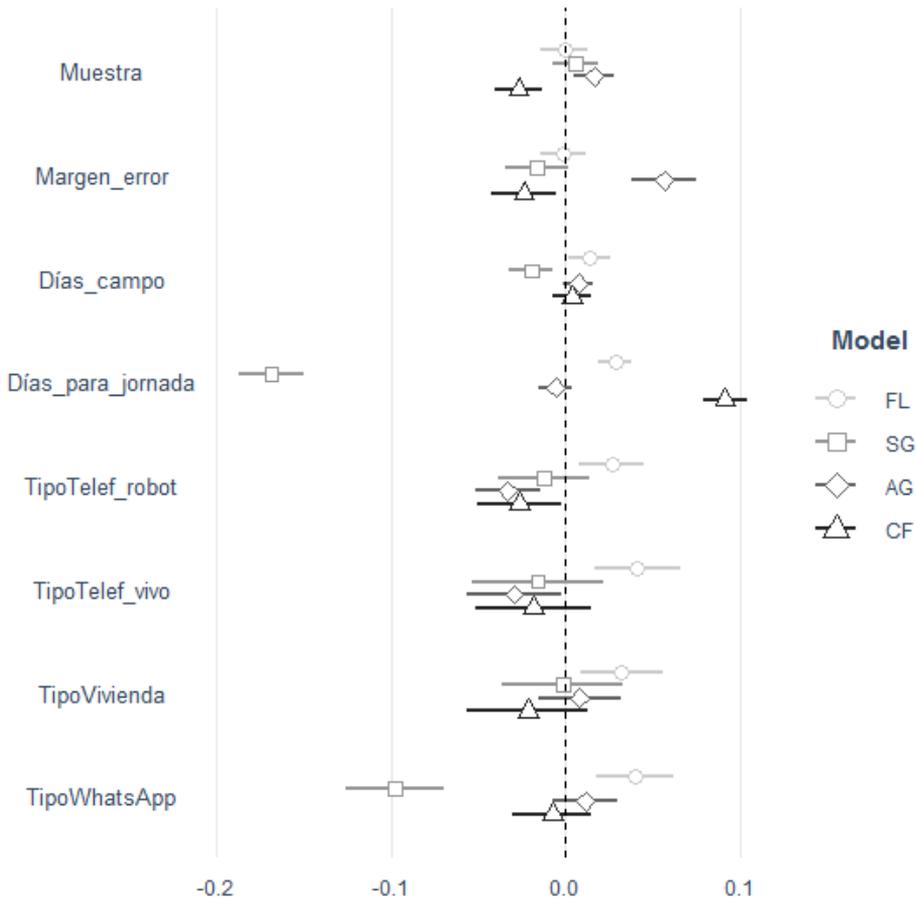
Fuente: Elaboración propia con base en el modelo interactivo de la Gráfica 3. Intervalos al 90 por ciento.

En el modelo interactivo se observa que el tamaño de muestra continúa negativamente asociado con el error promedio, además de que tres de los métodos de recolección a lo largo del tiempo de campaña están relacionados con M3: los dos tipos de levantamientos telefónicos y el de mensajería instantánea.

Para entender mejor el impacto de las interacciones, la Gráfica 2 muestra los valores pronosticados considerando el horizonte temporal, según cada método de recolección. Las encuestas en vivienda son las que más reducen su error promedio, por abajo de 2.5 puntos. Después se ubican los levantamientos telefónicos en vivo con 2.5 puntos de error y luego los métodos no probabilísticos, esto es, las encuestas telefónicas por robot y las realizadas a través de la red social Facebook, ambas entre 2.5 y 5 puntos de error promedio. Por último, restarían las de mensajería instantánea, vía WhatsApp, con un error promedio de entre 5 y 7.5 puntos.

Sobre las diferencias entre las mediciones por encuesta y el dato oficial, se estimaron modelos lineales con las mismas variables explicativas que los modelos donde M3 fungía de variable dependiente. Como se muestra en la Gráfica 3, se corrieron cuatro estimaciones, una por cada candidatura, las cuales sirven de modelo base. El margen de error y el tamaño de muestra se asocian positivamente con las mediciones sobre la intención de voto por de la Garza y negativamente por Flores, sugiriendo sobreestimación y cercanía al cero, respectivamente. Por su parte, la duración de los levantamientos y la distancia con el día de los comicios se asocian negativamente con García y positivamente con Flores y Larrazábal, sugiriendo mejores estimaciones para el primero que para las otras dos candidaturas.

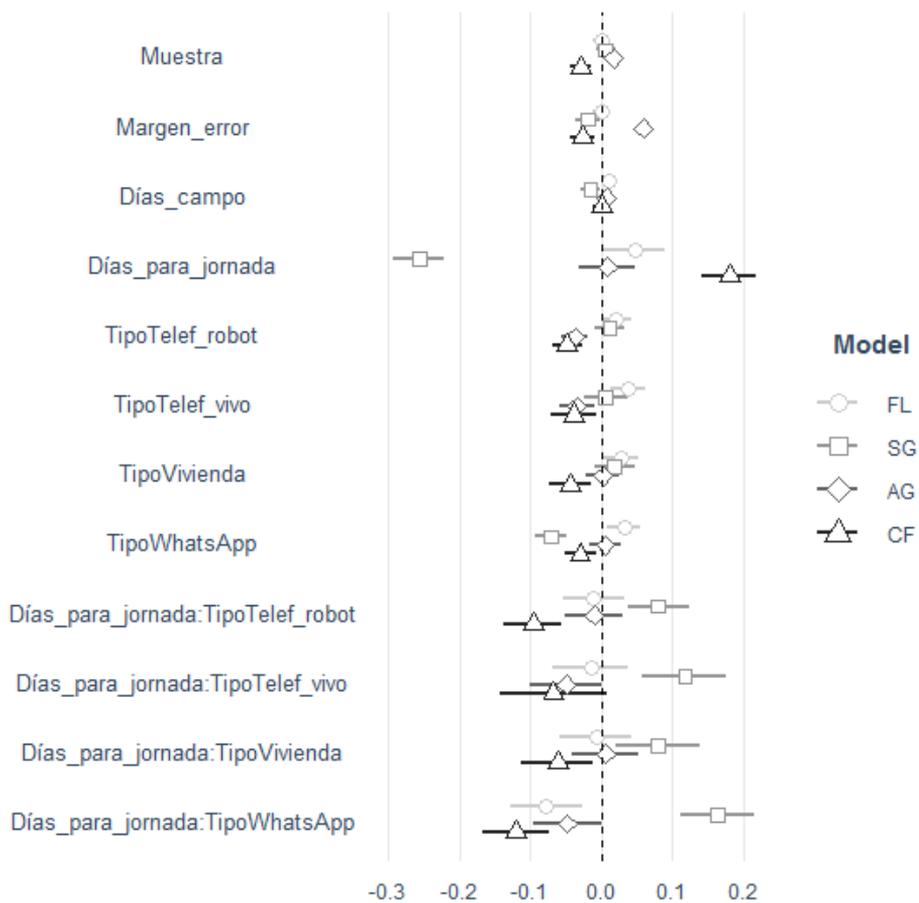
Gráfica 3. Determinantes del error de estimación por candidatura, modelos base



Fuente: Elaboración propia. Estimaciones lineales. Variable dependiente: error de estimación por candidatura. El método de recolección por Facebook sirve como categoría de referencia. 155 casos para todos los modelos. FL: R2: 0.23, prueba F: 6.7, error de la regresión 0.03. SG: R2: 0.69, prueba F: 44.57, error de la regresión: 0.06. AG: R2: 0.39, prueba F: 13.29, error de la regresión: 0.03. CF: R2: 0.60, prueba F: 29.77, error de la regresión: 0.04.

Respecto a la interacción entre los métodos de recolección y la distancia con la jornada electoral, la Gráfica 4 muestra los modelos lineales interactivos para cada candidatura, donde se encontraron asociaciones entre de la Garza, Larrazábal y Flores con las encuestas levantadas vía WhatsApp; de vivienda y telefónicas por robot con Flores, y de todos los métodos con García.

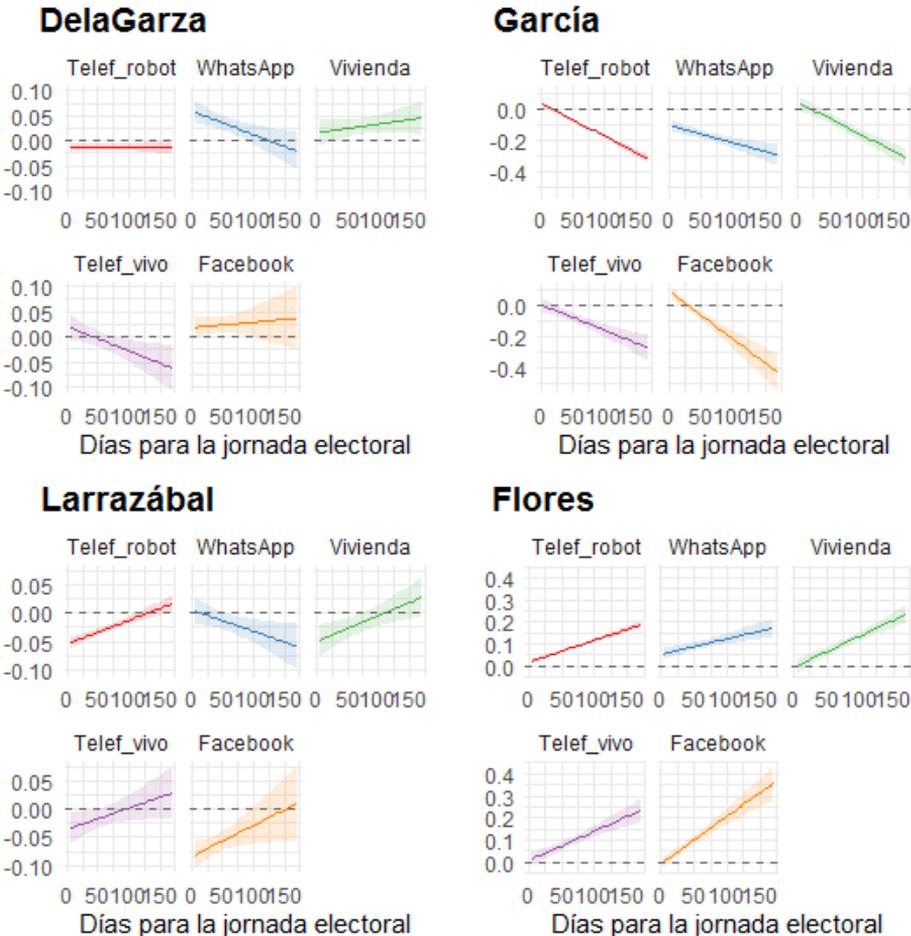
Gráfica 4. Determinantes del error de estimación por candidatura, modelos interactivos



Fuente: Elaboración propia. Estimaciones lineales. Variable dependiente: error de estimación por candidatura. El método de recolección por Facebook sirve como categoría de referencia. 155 casos para todos los modelos. FL: R2: 0.28, prueba F: 6.06, error de la regresión 0.03. SG: R2: 0.71, prueba F: 32.52, error de la regresión: 0.06. AG: R2: 0.42, prueba F: 10.17, error de la regresión: 0.03. CF: R2: 0.63, prueba F: 22.88, error de la regresión: 0.04.

Para entender mejor las interacciones, la Gráfica 5 presenta cada método de recolección por candidatura. En el primer panel se observan los métodos empleados para analizar a de la Garza, para quien las telefónicas por robot y Facebook no parecieron captar movimientos durante la campaña; aunque las primeras se ubicaron por abajo y las segundas por arriba de la línea cero. Las de WhatsApp sobreestimaron al candidato por 5 puntos en promedio, mientras que las telefónicas en vivo y las de vivienda lo hicieron por un poco menos de 2.5 puntos.

Gráfica 5. Predicción de probabilidades de error por candidatura, según días para la jornada y métodos



Fuente: Elaboración propia con base en el modelo interactivo de la Gráfica 6. Intervalos al 90 por ciento.

En el segundo panel se muestra el caso de García, donde las telefónicas en vivo terminaron prácticamente en la línea cero, después siguieron las de vivienda, junto con las de robot. En cambio, las de Facebook le sobrestimaron aproximadamente por 8 puntos, mientras que las de WhatsApp le subestimaron por un 10 por ciento. Por su parte, en el tercer panel se muestra la candidatura de Larrazábal, donde cuatro métodos le subestimaron: Facebook por más de 7.5 puntos, robot y vivienda por cinco puntos y telefónicas en vivo por 3. En contraste, las levantadas por WhatsApp se ubicaron en la línea cero. Finalmente, en el último panel se presenta la candidatura de Flores, donde vivienda, telefónicas en vivo y Facebook se ubicaron en la línea cero, seguidas por las de robot y después por las de WhatsApp, aunque estas últimas tuvieron 5 puntos de sobreestimación.

En síntesis, puede decirse que, tanto las encuestas en vivienda como las telefónicas en vivo, tuvieron un desempeño dentro de los márgenes deseables, mientras que las de robot y las de Facebook, aunque se acercaron en algunas candidaturas a la línea cero, no mostraron movimientos para el abanderado de la coalición PRI-PRD, pero sí para el resto. Finalmente, las de WhatsApp solo se acercaron a la línea cero en una candidatura, la presentada por el PAN, pero se mantuvieron alejadas en los casos restantes, incluso fuera de los márgenes de error deseables.

Conclusiones

Por más útil que resulte como caso de estudio, una elección en un solo estado del país difícilmente permite extraer conclusiones generales. Sin embargo, sí posibilita hacer un análisis de la evidencia compilada. En general, al explorar qué variables se asocian con el error promedio, se encontró que, conforme se acerca la fecha de los comicios, dicho error se reduce en todos los métodos de recolección. En particular, los probabilísticos presentaron niveles menores de error que los métodos no probabilísticos. Los errores de estimación se midieron a través de dos indicadores: el primero, conocido como M3, que es el error promedio y, el segundo, calculado como la diferencia entre la estimación por encuesta y el dato oficial para cada una de las cuatro candidaturas. Se empleó una compilación propia de un poco más de 150 mediciones a lo largo de cinco meses y medio, es decir, a partir de que fueron conocidas las candidaturas estudiadas.

Al comparar los resultados entre el error promedio y el error por candidatura, se encontró que las mediciones en vivienda y las telefónicas en vivo fueron de las más acertadas, pues se mantuvieron dentro de un margen de error aceptable. Por su parte, las telefónicas por robot, si bien se mantuvieron dentro del margen en tres casos, subestimaron a un candidato por cinco puntos. Respecto a las encuestas levantadas por Facebook, aunque en promedio se mantuvieron en el margen de error deseable, al revisar las candidaturas en lo individual, sub y sobrestimaron a un par de candidatos fuera de dicho margen. Finalmente, los levantamientos por WhatsApp se alejaron del margen promedio deseable, pues en tres de las cuatro

candidaturas excedieron los 5 puntos, subestimando al ganador y sobrestimando al segundo y al cuarto lugar.

Si bien se verificaron las hipótesis planteadas, esto es, que las encuestas probabilísticas redituaron mejores estimaciones que las no probabilísticas, todos los métodos de recolección se acercaron a la línea cero, es decir, redujeron los errores de estimación conforme se acercaba el día de la jornada electoral. Ello sin duda es una buena noticia para la pluralidad metodológica, pero hay que advertir que los métodos no probabilísticos aún requieren mejorar sus técnicas de medición, en particular, las administradas por mensajería instantánea.

Respecto a las características metodológicas, cabe precisar que, en algunos casos, se observó lo que la teoría predice, ya que una muestra mayor, más días en campo y un margen de error menor tendieron a reducir el error de estimación. Sin embargo, estas relaciones no siempre se sostuvieron para cada candidatura, por lo que quedará para futuras investigaciones indagar estas divergencias. Dos de las posibles razones por las cuales las encuestas a nivel estatal con cambios drásticos de preferencias se convierten en un reto mayúsculo para las casas encuestadoras son la alta volatilidad y el reto de estimar un blanco en movimiento, en vez de uno fijo.

La primera situación ya se había observado en la elección de Nuevo León en 2015, en donde se generaron estimaciones con mayores errores promedio que los casos federales y con una distancia entre el primero y el segundo lugar de dos dígitos. La segunda situación, la alta volatilidad, podría ser como en las elecciones presidenciales de 2012 (Moreno, 2018): las nuevas encuestas dan cuenta de cambios en las preferencias, ocasionando que las estrategias de los partidos y de los equipos de campaña comiencen a modificarse, alterando las preferencias de algún segmento del electorado y, a su vez, las sucesivas mediciones. Si estas razones resultan plausibles, investigaciones futuras podrían compilar mediciones en numerosas entidades federativas e indagar si los casos de cambios drásticos en las preferencias dan pie a cambios como los esbozados.

Finalmente, se podría concluir que la cercanía con la jornada electoral se asocia con un menor error promedio en las estimaciones para todos los métodos de recolección estudiados, pero que este varía cuando se estudia cada candidatura en lo individual, por lo que se requiere de un análisis puntual de los casos en los que algunos métodos fueron más exitosos que otros. En síntesis, si bien es cierto que el pluralismo metodológico debe ser siempre bienvenido, los métodos no probabilísticos requieren afinar sus metodologías para ofrecer estimaciones más certeras, no solo en la métrica de evaluación general o promedio, sino también en la mayoría de las métricas puntuales, esto es, por candidatura. Lo anterior resulta necesario para despejar dudas respecto a la utilidad de métodos novedosos, los cuales aún deben perfeccionarse.

Referencias

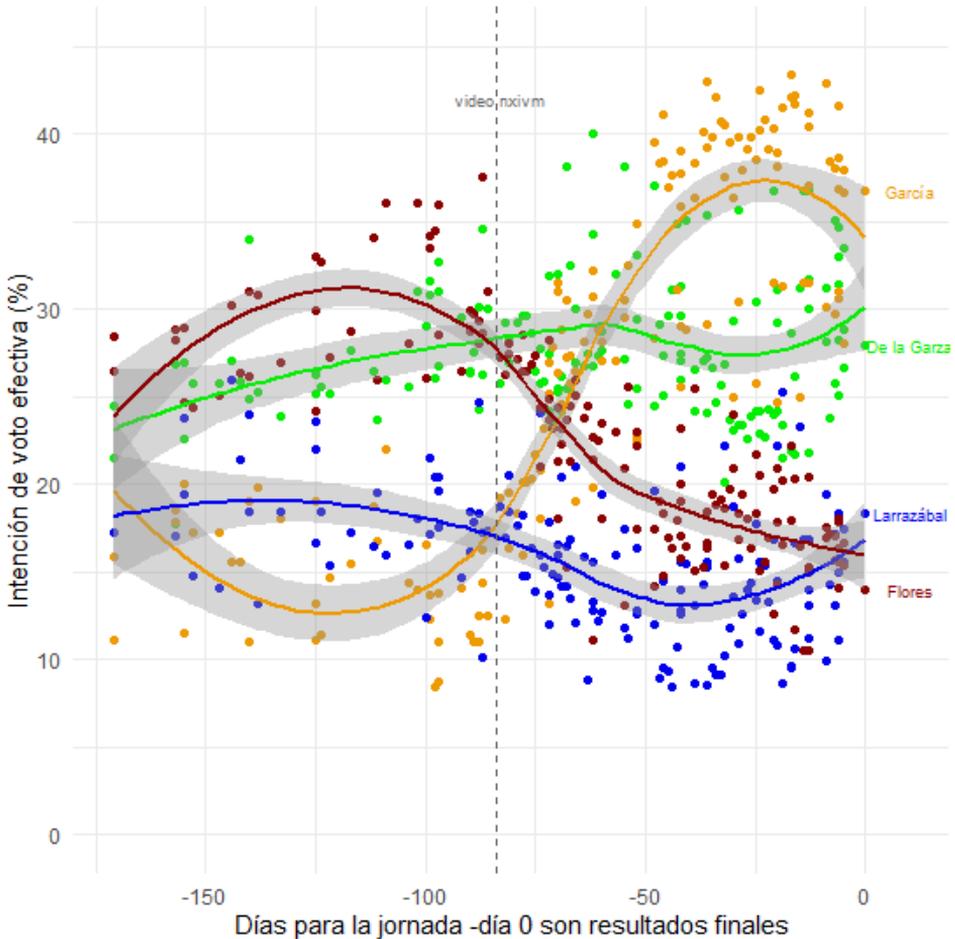
- Antoun, C., Couper, Mick y Conrad, Frederick. (2017). Effects of Mobile versus PC Web on Survey Response Quality: A Crossover Experiment in a Probability Web Panel. *Public Opinion Quarterly*, 81(S1), 280-306. <https://doi.org/10.1093/poq/nfw088>
- Arzheimer, K. y Evans, J. (2014). A New Multinomial Accuracy Measure for Polling Bias. *Political Analysis*, 22(1), 31-44. <https://doi.org/10.1093/pan/mpt012>
- Cantú, F., Hoyo, V. y Morales, M. (2016). The Utility of Unpacking Survey Bias in Multiparty Elections: Mexican Polling Firms in the 2006 and 2012 Presidential Elections. *International Journal of Public Opinion Research*, 28(1), 96-116. <https://doi.org/10.1093/ijpor/edv004>
- Castorena, O., Lupu, N., Schade, M. y Zechmeister, E. (2023). Online Surveys in Latin America. *PS: Political Science & Politics*, 56(2), 273-280. <https://doi.org/10.1017/S1049096522001287>
- Cornesse, C., Blom, A., Dutwin, D., Krosnick, J., De Leeuw, E., Legleye, S., Pasek, S., Pennay, D., Phillips, B., Sakshaug, J., Struminskaya, B. y Wenz, A. (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. *Journal of Survey Statistics and Methodology*, 8(1), 4-36. <https://doi.org/10.1093/jssam/smz041>
- CEE. (2021a). *Obligaciones de quienes publiquen, soliciten u ordenen encuestas por muestreo, sondeo de opinión, encuestas de salida y conteos rápidos, Proceso Electoral 2020-2021*. Portal anterior del Instituto Estatal Electoral y de Participación Ciudadana de Nuevo León. <https://www.ceenl.mx/pe2020/informes/>
- CEE. (2021b). *Cómputos 2021. Elecciones Estatales de Nuevo León, Gubernatura*. Comisión Estatal Electoral de Nuevo León. <https://computo2021dc.ceenl.mx/R03E.htm>
- De la Garza, D. (2022). La cuarta alternancia en el estado de Nuevo León desde la transición a la democracia: redes sociales, polarización, identidad regional y rebelión contra el centralismo. *Revista Mexicana de Opinión Pública*, 32(1), 135-149. <https://doi.org/10.22201/fcpys.24484911e.2022.32.80652>
- De la Peña, R. (2020). Las encuestas en la elección presidencial de México, 2018. *Revista Mexicana de Opinión Pública*, 28(1), 116-145. <https://doi.org/10.22201/fcpys.24484911e.2020.28.66376>
- Díaz Domínguez, A. (en prensa). Encuestas y partidos políticos en México: 30 años de regulaciones. En S. Loaeza y T. Payán (comps.). *Partidos, procesos y prácticas. La terca pluralidad política de la democracia mexicana*. El Colegio de México.
- Durand, C. y Johnson, T. (2021). What About Modes? Differences Between Modes in the 21st Century's Electoral Polls Across Four Countries. *Public Opinion Quarterly*, 85(1), 183-222. <https://doi.org/10.1093/poq/nfab006>
- Galindo, J. (1 de junio de 2021). Encuestas: Samuel García llega a la recta final por delante en Nuevo León. *El País*. <https://elpais.com/mexico/2021-06-02/encuestas-samuel-garcia-llega-a-la-recta-final-por-delante-en-nuevo-leon.html>
- Fienberg, S. y Hoaglin, D. (2006). *Selected Papers of Frederick Mosteller*. Springer. <https://doi.org/10.1007/978-0-387-44956-2>
- Hartman, E. y Levin, I. (2019). Accounting for Complex Survey Designs: Strategies for Post-stratification and Weighting of Internet Surveys. En E. Suhay et al. (comps.). *The Oxford*

- Handbook of Electoral Persuasion*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190860806.013.29>
- Jennings, W. y Wlezien, C. (2018). Election Polling Errors across Time and Space. *Nature Human Behaviour*, (2), 276-283. <https://doi.org/10.1038/s41562-018-0315-6>
- Kennedy, C., Blumenthal, M., Clement, S., Clinton, J., Durand, C., Franklin, C., McGeeney, K., Miringoff, L., Olson, K., Rivers, D., Saad, L., Witt, E. y Wlezien, C. (2018). An Evaluation of 2016 Election Polls in the United States. Report commissioned by the American Association of Public Opinion Research. *Public Opinion Quarterly*, 82(1), 1-33. <https://doi.org/10.1093/poq/nfx047>
- Kott, P. y Liao, D. (2015). One Step or Two? Calibration Weighting from a Complete List Frame with Nonresponse. *Survey Methodology*, 41(1), 165-181. <https://www150.statcan.gc.ca/n1/pub/12-001-x/2015001/article/14172-eng.htm>
- Martin, E., Traugott, M. y Kennedy, C. (2005). A Review and Proposal for a New Measure of Poll Accuracy. *Public Opinion Quarterly*, 69(3), 342-369. <https://doi.org/10.1093/poq/nfi044>
- Moreno, A., Aguilar, R. y Romero, V. (2014). Estimaciones de encuestas preelectorales en México: en busca de las principales fuentes de error. *Revista Latinoamericana de Opinión Pública*, (4), 49-93. <https://doi.org/10.14201/rlp.22299>
- Moreno, A. (2018). *El cambio electoral. Votantes, encuestas y democracia en México*. Fondo de Cultura Económica.
- Mosteller, F., Hyman, H., McCarthy, P., Marks, E. y Truman, D. (1949). *The Preelection Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. Boletín, Social Science Research Council, (60).
- Panagopoulos, C., Endres, K. y Weinschenk, A. (2018). Preelection Poll Accuracy and Bias in the 2016 U.S. General Elections. *Journal of Elections, Public Opinion and Parties*, 28(2), 157-172. <https://doi.org/10.1080/17457289.2018.1441850>
- Rich, R., Brians, C. L., Manheim, J. y Willnat, L. (2018). *Empirical Political Analysis. Quantitative and Qualitative Research Methods*. Routledge.
- Romero, V. (2012). Notas para la evaluación de las encuestas preelectorales. Las elecciones para gobernador de 2010 en México. *Política y Gobierno*, XIX(1), 101-124. <http://www.politicaygobierno.cide.edu/index.php/pyg/article/view/153>
- Saris, W. y Gallhofer, I. (2014). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Wiley. <https://doi.org/10.1002/9780470165195>
- Sohlberg, J., Gilljam, M. y Martinsson, J. (2017). Determinants of Polling Accuracy: The Effect of opt-in Internet Surveys. *Journal of Elections, Public Opinion and Parties*, 27(4), 433-447. <http://dx.doi.org/10.1080/17457289.2017.1300588>
- Stier, S., Breuer, J., Siegers, P. y Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, 38(5), 503-516. <https://doi.org/10.1177/0894439319843669>
- Valle-Jones, D. (2012). *Final Poll of Polls 2012*. Diego Valle-Jones [Blog]. <https://blog.diegovalle.net/2012/06/final-poll-of-polls.html>

Métodos de recolección y precisión de encuestas. El caso de Nuevo León, México en 2021

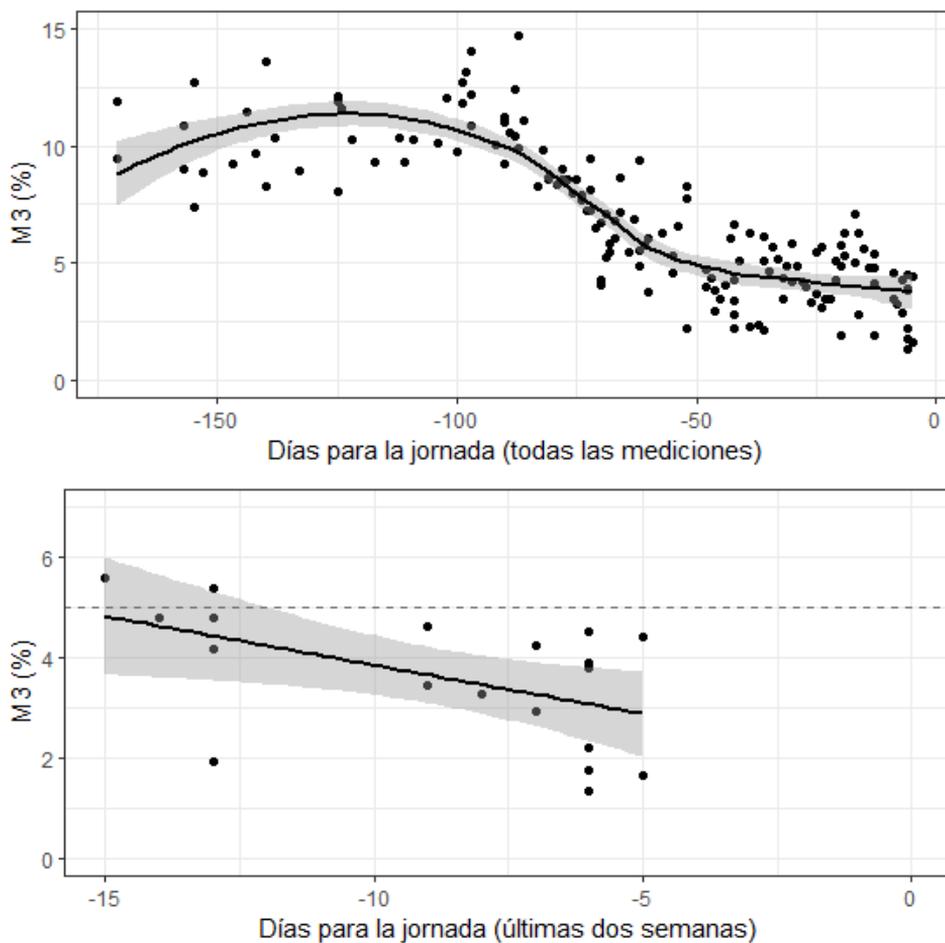
Anexo en línea

Gráfica A1. Las campañas a la gubernatura en Nuevo León 2021 según las encuestas



Fuente: Elaboración propia con base en los datos de la tabla 1. 155 mediciones de 17 casas encuestadoras. Líneas e intervalos al 95 por ciento son estimaciones de regresiones locales. Resultados oficiales colocados en día cero.

Gráfica A2. Días para la jornada electoral y error promedio, Nuevo León 2021



Fuente: Elaboración propia con base en los datos de la tabla 1. Panel superior muestra un horizonte temporal de cinco meses y medio. Panel inferior muestra dos semanas previas al día de la jornada electoral. Eje vertical muestra el error promedio (M3).