

¿POR QUÉ Y PARA QUÉ UNA ENCUESTA DE SEGUIMIENTO DIARIO?

Ricardo de la Peña.

Presidente Ejecutivo de

ISA Investigaciones Sociales Aplicadas®

Si bien fue en 1988 cuando aparecieron por vez primera en medios encuestas nacionales previas a elecciones presidenciales, fue seis años después cuando se inició la publicación de series de mediciones nacionales por encuesta, realizadas por empresas investigadoras especializadas y patrocinadas por medios de comunicación. Ello era la natural respuesta a dos fenómenos: la emergencia de condiciones propiamente democráticas en la competencia electoral en México y el desarrollo de tecnologías que permitían la realización de estudios por muestreo en tiempos y con costos relativamente reducidos.

A partir de entonces, en cada elección presidencial se dispuso de diversas series de encuestas nacionales, producidas por distintas casas encuestadoras para diferentes medios, con periodicidades próximas a lo mensual, aunque mayor frecuencia hacia el cierre del proceso. En la elección anterior, fueron más de setenta encuestas nacionales en vivienda hechas públicas y tres de cada cuatro de ellas conformaron series, con la consecuente homogeneidad en métodos y alcances en sus estimaciones.

Para la elección presidencial de 2012 era esperable que se ahondara en el esfuerzo por dar cuenta de las preferencias de los ciudadanos y sus cambios en el tiempo. La experiencia de 2006 permitía suponer que estaban dadas las condiciones para que medios y encuestadoras acordaran llevar adelante series con mediciones más frecuentes. Y eso fue lo que ocurrió: además de las series mensuales a las que estamos acostumbrados, una casa encuestadora programó un seguimiento semanal, que luego se acompañó por otro ejercicio con igual frecuencia realizado por otra marca. Y un grupo editorial resolvió llevar adelante un seguimiento diario mediante encuesta: Milenio, contratando para ello a la firma GEA-ISA.

El objetivo de operar una encuesta nacional de seguimiento diario sobre preferencias para la elección presidencial de 2012 fue una inquietud atendida por esta casa encuestadora desde un año antes de los comicios. Sabíamos de entrada que era inviable técnica y financieramente realizar dicho ejercicio recurriendo a tecnologías y metodologías convencionalmente empleadas para el desarrollo de estudios por encuesta en una medición aislada. De hecho, se contaba con dos elementos para comenzar a definir el modelo: uno, lo realizado en otras naciones y por otras empresas en el pasado; y dos, nuestra propia experiencia en operativos de esta naturaleza.

Desde luego, para el caso mexicano había que tomar en cuenta un aspecto adicional: dada la reducida cobertura de la telefonía en domicilios, este método de aproximación estaba descartado si lo que se pretendía era disponer de estimadores relativos al universo de los electores. Así, el único método de aproximación posible válido era y es la entrevista personal en vivienda.

Conocíamos así los métodos empleados para las encuestas de seguimiento, mediante estimaciones de medias móviles para períodos de tres días, lo que reducía el número de casos a observar a la tercera parte. Además, adoptando sistemas de rotación diaria de las unidades en muestra y no sorteos independientes para cada jornada, se logra una relativa parsimonia en el cambio de las estimaciones, a la vez que se alcanza una sensible reducción de costos por estancias y traslados, siempre y cuando se dispusiera de un grupo de trabajo distribuido de manera adecuada para atender la demanda del proyecto.

Para la definición de la estrategia de operación en una encuesta de seguimiento para 2012, había que tomar en cuenta asuntos operativos adicionales. Dado que es imposible lograr un acopio manual de cuestionarios en varios puntos de entrevista por una misma persona, tenía que suprimirse toda intermediación y adoptar mecanismos orientados a la trasmisión directa y en tiempo real de entrevistas, pero sin menoscabo del control de calidad que garantice la efectiva y adecuada aplicación.

Ello sólo era posible arribando al empleo de tecnologías de comunicación en línea mediante teléfonos inteligentes. Pero, para el control de calidad se requería disponer de plataformas de control de ubicaciones geográficas de operadores, por lo que se solicitó y obtuvo del IFE la información para la referencia geográfica de las secciones electorales. Así, todo caso que se trasmite es checado de manera automática por el sistema, para cotejar que las coordenadas de la trasmisión correspondan con el perímetro de operación elegido; caso contrario, el caso se invalida. Adicionalmente, se dispone del domicilio y descripción de la vivienda, por lo que se puede consultar de inmediato su existencia en cartografía y corroborar que la descripción coincida con archivos públicos de imágenes. Asimismo, al acopiar números de teléfonos para el posterior contacto con un entrevistado, se pueden efectuar rutinas de verificación mediante llamadas. Luego, si bien se recurre a procedimientos convencionales de supervisión en campo, mediante itinerarios que permiten la cobertura paulatina y sin previo aviso de los operadores, el control de calidad de los estudios no tiene por qué reducirse a estos procedimientos.

Con base en estas decisiones, en el tercer trimestre de 2011 se hicieron las estimaciones sobre los requerimientos tecnológicos, se hizo un estudio del mercado de oferentes de servicios de trasmisión de datos, conociendo y evaluando sus plataformas y mecanismos de comunicación; con base en dicho estudio, se contrató los servicios

con un proveedor que aportaba condiciones tecnológicas y económicas idóneas para la operación.

Esto permitió disponer de una primera aproximación a los costos involucrados en el proyecto. Con base en ello, en el último trimestre del año pasado se iniciaron acercamientos con potenciales patrocinadores, para conocer la posibilidad de trabajar conjuntamente un estudio de esta naturaleza orientado a su difusión pública, lo que concluiría en el convenio celebrado con el Grupo Milenio, para que éste difundiera los resultados y se responsabilizara cabalmente del patrocinio.

En febrero pasado, se llevó a cabo el diseño definitivo y selección de la muestra para el estudio nacional, incluyendo reemplazos potencialmente necesarios y criterios para enfrentar las dificultades esperables en este tipo de operativos. Asimismo, se avanzó en la definición de los contenidos e instrumentos de recuperación de datos, considerando la información mínima deseable y la limitación de tiempos de entrevista por el recurso a nuevas opciones tecnológicas para el acopio y trasmisión de datos.

Además, para la puesta en punto del proyecto, se realizaron diversos ejercicios con la intención de probar la tecnología, medir tiempos y efectividad de la toma y trasmisión de datos, probar la plataforma de recuperación que alimenta la base de datos y probar rutinas de traslado de la información a programas convencionales. Asimismo, se efectuaron mediciones para cotejar el eventual impacto de la adopción de procedimientos de entrevista asistida por computadora versus los procedimientos tradicionales con papel y lápiz.

Igualmente, con base en la distribución geográfica de la muestra elegida, se seleccionaron las sedes en que se requería disponer de personal calificado para operar el estudio y se efectuó el debido reclutamiento y capacitación. El perfil de operadores es el siguiente: la mitad son hombres y la otra mitad mujeres; su edad va de los veinte a los sesenta años, con una media próxima a los treinta años; dos tercios cuentan con estudios superiores y el resto de nivel medio.

En paralelo, se definieron algunos principios básicos de carácter metodológico para la operación del proyecto: tamaño deseable de la muestra y dispersión requerida. Conforme estas estimaciones, se decidió que se requeriría disponer de una muestra de 48 secciones electorales elegidas con probabilidad proporcional al tamaño, definido por el número de electores registrados en el Listado Nominal.

Como segunda etapa, se elegirían ocho viviendas por sección, mediante recorrido aleatorio, con arranque a partir del domicilio de la casilla básica en la elección federal pasada e intervalo constante por cada sección en muestra. Para la selección del informante, se trataría de tomar datos a la persona que, formando parte de la población objetivo, celebrara el próximo cumpleaños de entre las presentes al momento de la entrevista.

En caso de no poder realizar la entrevista en el domicilio elegido o a la persona seleccionada, se procedería a entrevistar en una vivienda contigua, hasta su realización, llevando control de la tasa de rechazo, asumiendo que ésta correspondería al cociente de los casos de entrevista completa transmitidas automáticamente entre contactos realizados para lograr dichas entrevistas, incluyendo entrevistas completas, incompletas, rechazos y casos en que se hubiera logrado el contacto, pero en los que en la vivienda seleccionada no residieran personas con las características demandadas en la población a observar. Por razones prácticas, se excluirían del cálculo los casos en que no se lograra el contacto o no existiera información que permitiera determinar la elegibilidad de los residentes en una vivienda.

Las entrevistas serían personales (cara a cara), en viviendas particulares, aplicando de manera asistida por computadora un cuestionario estructurado, cuyo contenido estaría exclusivamente vinculado a este ejercicio. La transmisión de datos sería en línea para su compilación; en caso de no poder transmitirse de manera automática los datos por cobertura o cualquier otro motivo, se recuperarían en papel, para su posterior transmisión.

Para fines del reporte público de estimaciones sobre preferencias electorales, se agregarían los 1152 casos tomados durante tres días consecutivos en 144 distintas secciones electorales, lo que permitiría cumplir los requerimientos para disponer de datos con un margen de error teórico de ± 3 por ciento al 95 por ciento de confianza.

El procedimiento para el cálculo de proporciones por candidato adoptaría dos modalidades: la estimación a partir del total de observaciones, incluyendo los casos en que el entrevistado no se definiera por un candidato o partido por cualquier razón; y la estimación excluyendo los casos en que el entrevistado no se definiera por alguien, mediante la simple exclusión de los casos no definidos para fines de cálculo.

Los resultados presentados no serían en ningún caso frecuencias simples, sino estimaciones basadas en el empleo de factores de expansión muestral, calculados como el inverso de la probabilidad de selección de cada individuo en muestra, ajustados para hacer acorde la distribución diaria de casos por sexo y edad con la reportada en el Listado Nominal de Electores.

Todo ello respondía a un doble requerimiento, propio de las encuestas de seguimiento cotidiano: buscar la estabilidad básica, pero una sensibilidad a los cambios en la coyuntura. De lo que se trata en estos ejercicios es de mostrar las variaciones en el día al día, en el marco de una distribución básica que puede ser sostenida a lo largo del tiempo.

Esto viene a cuento porque en México ha prevalecido un mito que tiende a descalificar la variabilidad en las mediciones por encuesta, en aras de la afirmación de una parsimonia que supuestamente reflejaría mayor calidad en los ejercicios.

Y sin embargo, las preferencias se mueven. Eso es al menos lo que estamos detectando en este proceso electoral 2012, donde un reparto básicamente estable no deja de estar sacudido por fluctuaciones de corto plazo.

¿Qué historia nos está contando nuestra encuesta de seguimiento? Uno podría privilegiar la lectura semanal que muestra cambios menores a lo largo del tiempo y entre corte y corte. Pero también podría uno atender los cortes diarios y ver que existe una fluctuación cotidiana que se esconde detrás de una serie discreta de mediciones semanales.

¿A qué se deben estas fluctuaciones? De la revisión de los datos acopiados, se descubren que varias posibles fuentes de variación no son su causa: no son producto de cambios en la no respuesta a la pregunta sobre preferencias, aunque sí se detecta una esperada sensibilidad en los niveles de indefinición, menores ahora que antes de campañas. Tampoco pareciera ser producto de tomas de datos en días específicos: claro que las tomas diarias varían más que las medias, pero dentro de un margen esperado, dada la cantidad de casos observados por jornada. No se detectan tampoco patrones de estacionalidad según día de la semana u horario de entrevista, lo que reflejaría que los métodos de selección de entrevistado y ponderación cotidiana de casos en muestra están permitiendo tener poblaciones comparables día con día y en los distintos horarios.

Algunos hechos que sí se observan son: los movimientos que se registran tienden a conformar bucles, sustentando la idea de que pudiera haber una tendencia regresiva a una media. Las fluctuaciones se dan dentro de una banda definida que no pareciera haber cambiado a lo largo de la serie, salvo para el componente menor, aunque ello no supone necesariamente la homogeneidad en los datos, más sí la inexistencia hasta el momento de una tendencia clara.

Aunque hay variaciones día con día, éstas no resultan mayores, toda vez que la práctica totalidad de los puntos caen dentro del intervalo estimado la jornada anterior. Puede afirmarse que las fluctuaciones observadas en este ejercicio no son de ninguna manera anómalas o peculiares cuando se les coteja con las variaciones observadas en ejercicios similares. El giro medio se ubica por debajo de dos puntos. Este nivel de variación no resulta sorprendente si se compara con el ritmo de un punto y medio medido por Gallup para el proceso electoral de 2008 en Estados Unidos, donde existe un sistema bipartidista y no con tres componentes importantes, como el nuestro.

Lo que es más: el cambio por candidato se sitúa en torno a 1.2 puntos diarios. Si lo comparamos con las variaciones en el nivel de aprobación presidencial registradas en seguimientos diarios en Estados Unidos durante el mismo período, se detecta una

sorprendente similitud: Rasmussen mide 1.1 puntos de cambio al día y Gallup alcanza 1.3 puntos de variación.

Hasta el momento, el seguimiento diario por nuestra encuesta ha registrado fluctuaciones que se han correspondido con los cambios observados por otras series discretas de encuestas. Salvo contadas excepciones, los intervalos de las mediciones públicas de otras casas encuestadoras se intersecan con los intervalos registrados por la encuesta de seguimiento en los mismos días. Lo que es más: los valores promedios por candidato medidos por GEA-ISA para Milenio durante la encuesta de seguimiento se ubica prácticamente en los mismos niveles que el promedio de las encuestas publicadas en el mismo período.

Así, la encuesta Milenio/GEA-ISA estaría aportando al público un seguimiento cotidiano prácticamente en tiempo real de las fluctuaciones coyunturales en las preferencias, que tienden a ser luego confirmadas por otras firmas. Esto fue notable luego del primer debate entre candidatos, donde nuestra encuesta de seguimiento fue referente inmediato que dio cuenta del impacto del evento prácticamente en tiempo real, con efectos que luego fueron corroborados por las mediciones de otras casas encuestadoras.

De ser cierto lo anterior, sería de esperarse que las variaciones medidas por las series discretas oculten movimientos intermedios. Habría así, además de tendencias de medio plazo medidas desde antaño, cambios de corto plazo e incluso fluctuaciones de corta duración, pero posible registro, que estarían siendo documentadas por vez primera mediante este ejercicio de seguimiento cotidiano.

Adentrados ya en la vigesimoprimer centuria, es importante adoptar métodos de medición que respondan a una nueva realidad, donde los tiempos en que el elector demanda información son cada vez más inmediatos. Otorgar datos sobre preferencias de manera casi inmediata es luego no sólo posible, sino adecuado. Lo que sería de esperar es que en posteriores ocasiones se sumen oferentes que permitan el contraste, todo ello desde la perspectiva de que en una democracia moderna los ciudadanos deben estar informados y que la pluralidad de fuentes es la mejor opción.