

Nota Metodológica #002

Evaluación y mejoramiento de la calidad de las entrevistas en el Barómetro de las Américas 2016/17

Mollie J. Cohen y Sebastián Larrea

LAPOP, Vanderbilt University

mollie.j.cohen@vanderbilt.edu

sebastian.c.larrea@vanderbilt.edu

6 de febrero de 2018

Hallazgos principales:

- El protocolo de auditoría humana de LAPOP mejora la calidad de la entrevista en el curso del trabajo de campo.
- La fabricación de datos está presente en una proporción muy pequeña de entrevistas—el 7 % de las canceladas y menos del 2 % de todas las entrevistas realizadas.
- La gran mayoría de los errores en las encuestas del Barómetro de las Américas están relacionados con la lectura completa y correcta de las preguntas. Incluso estos errores disminuyen a lo largo del trabajo de campo al utilizar protocolos de auditoría humana.



Avances metodológicos recientes han permitido a los investigadores identificar posibles falsificaciones y fabricaciones de datos en bases de datos de encuestas (e.g., Kuriakose y Robbins 2016). Históricamente, sin embargo, los investigadores por muestreo no han sido capaces de identificar y detener la fabricación de datos en tiempo real, o de proveer apoyo adicional y entrenamiento a los entrevistadores que lo necesitan mientras el trabajo de campo se está llevando a cabo. El sistema de control de calidad de LAPOP, particularmente la implementación de auditorías casi en tiempo real, representa un avance sustancial hacia este objetivo. Esta Nota Metodológica describe las prácticas¹ de auditoria humana de LAPOP para asegurar la calidad de los datos (un protocolo que depende en gran medida del módulo de control de calidad–el “QuAC²”–programado en nuestro software de recopilación de datos), y describe los errores que documentamos y abordamos durante el trabajo de campo en diez países incluidos en el estudio 2016/17 del Barómetro de las Américas. La gran mayoría de estos errores, dentro del 14 % de entrevistas que eventualmente fueron canceladas, son problemas relacionados con la lectura correcta y completa de las preguntas; es decir, la mayoría de los errores indican deficiencias en la calidad de las entrevistas, y no tienen que ver con la fabricación de datos. También demostramos que la implementación de la auditoría casi en tiempo real puede mejorar la calidad de las entrevistas a medida que progresá el trabajo de campo.

Auditoría en tiempo cuasi-real utilizando el QuAC

Para la ronda 2016/17 del Barómetro de las Américas, y como parte del protocolo FALCON[®], LAPOP implementó la captura silenciosa de imágenes de los entrevistadores, ubicaciones GPS y las grabaciones de audio de preguntas clave³. Estas prácticas permiten a los equipos que trabajan en los proyectos de LAPOP, evaluar la calidad de las entrevistas en tiempo real y proveer retroalimentación a los entrevistadores durante el trabajo de campo⁴. Durante el trabajo de campo, auditores en las oficinas centrales de LAPOP y el personal en nuestros equipos de campo comparan

las imágenes capturadas silenciosamente con la cámara frontal con las fotografías de los entrevistadores para garantizar que los entrevistadores en el campo hayan sido los mismos que fueron entrenados por el personal de LAPOP. Además, analizamos las coordenadas GPS de las entrevistas para asegurarnos de que cada entrevista se haya realizado en la ubicación asignada (y que la ubicación sea, de hecho, una residencia en lugar de un parque público o un centro comercial). Finalmente, escuchamos las grabaciones de audio para asegurarnos de que los entrevistadores hayan leído las preguntas de manera completa y correcta, sin interpretar las preguntas con sus propias palabras, ni que se hayan saltado preguntas o hayan influenciado en las respuestas de los entrevistados⁵.

Sobre la base de estas auditorias, asignamos a cada entrevista un puntaje de control de calidad. Los puntajes más altos indican problemas más graves, y nos negamos a aceptar entrevistas (es decir, requerimos su cancelación) con puntajes superiores a 20 puntos. En los diez países en que usamos la versión final del QuAC, cancelamos el 14.3 % de todas las entrevistas debido a una amplia gama de errores⁶.

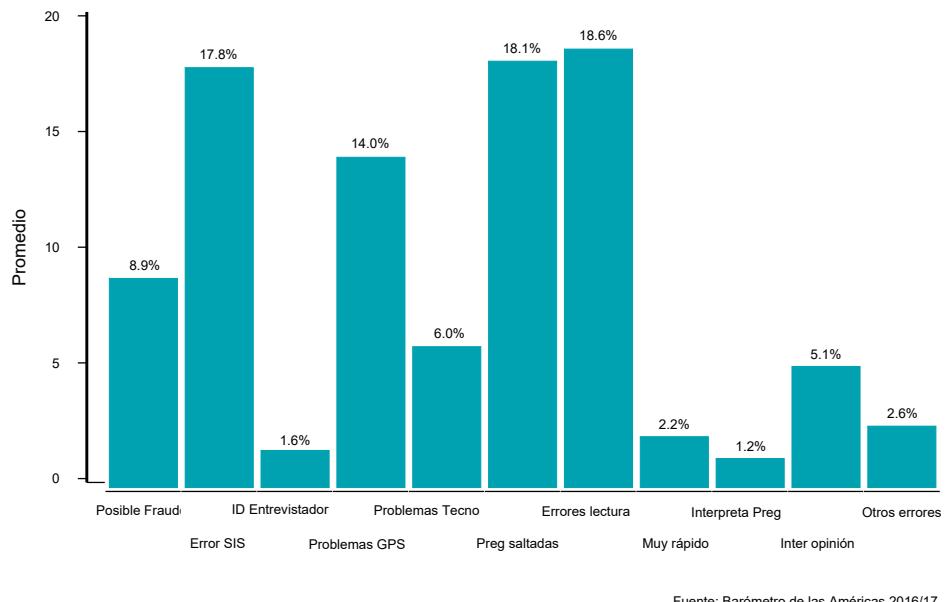
¿Cómo funciona el control de calidad, casi en tiempo real, en la práctica? En la primera semana de cada estudio, el personal audita todas las entrevistas^{7,8}. En cuestión de días, queda claro qué entrevistadores son sobresalientes (miembros del “Equipo A”), y qué entrevistadores necesitan orientación adicional para mejorar (miembros del “Equipo B”)⁹. Luego cambiamos nuestra estrategia, revisando el trabajo del Equipo A periódicamente, y concentrando la mayor parte de nuestros esfuerzos en los miembros del Equipo B. Dentro de las 48 horas posteriores a cada entrevista, enviamos a los equipos de trabajo de campo local información identificando entrevistas problemáticas y sugerencias para mejorar el trabajo de entrevistadores específicos. LAPOP puede, por lo tanto, proporcionar retroalimentación a los encuestadores casi en tiempo real¹⁰.

Típicamente, la retroalimentación que proveemos es general. “Por favor recuerde a Juan Pérez que lea las preguntas completa y correctamente”. “Con frecuencia, María López habla demasiado rápido como para que los encuestados la entiendan; recuérdelle hablar a un ritmo más lento”.

A menudo, identificamos entrevistas específicas que ejemplifican estos errores. Proveer este tipo de retroalimentación funciona: durante el trabajo de campo, encontramos menos entrevistas problemáticas ya que los encuestadores recuerdan seguir las mejores prácticas—leer todas las preguntas y opciones de respuesta tal y como están escritas, leer a un ritmo normal y leer la hoja de información del estudio (SIS por sus siglas en inglés)¹¹ sin errores.

La retroalimentación positiva también es importante. Creemos que los miembros de los equipos A y B merecen reconocimiento por su buen trabajo, por lo que hacemos nuestro mejor esfuerzo por brindar buenas noticias a los equipos locales. Estos comentarios también son generales. Por ejemplo: “Las entrevistas de Pablo Flores han mejorado notablemente esta semana; él está tomando en cuenta nuestros comentarios sobre la lectura completa y correcta de todas las preguntas. Por favor comuníquele que hemos notado su mejoría” o “Juana Díaz ha hecho un trabajo sobresaliente como se muestra en la siguiente entrevista. Por favor comuníquele que siga con el buen trabajo”.

La mayoría de problemas que encontramos en las entrevistas son errores de lectura. El Gráfico 1 muestra la distribución de errores en las entrevistas que posteriormente fueron canceladas¹². Las entrevistas pueden tener más de un error, por lo que estos valores no suman el 100 %. Los errores de lectura, los problemas con el GPS y los errores al leer la "SIS" fueron mucho más comunes que los errores relacionados con la fabricación de datos (que ocurrieron solamente en el 8.9 % del 14.3 % de las entrevistas que finalmente se cancelaron). Esto sugiere que la mayoría de errores en nuestras encuestas son el resultado de que los entrevistadores olvidan o ignoran las lecciones fundamentales de su capacitación. Si recordarles las mejores prácticas a los entrevistadores mejora su trabajo, nuestro mecanismo de retroalimentación debería resultar en una disminución de los puntajes de control de calidad (es decir, mejores niveles de calidad de las entrevistas) y menos cancelaciones en el transcurso del trabajo de campo, ya que los entrevistadores gradualmente dejan de cometer estos errores. A continuación analizamos estos datos.



Fuente: Barómetro de las Américas 2016/17

Gráfico 1: Errores encontrados en entrevistas canceladas

Puntajes y cancelaciones del QuAC a lo largo del tiempo

¿Cómo cambia la calidad de las entrevistas en el transcurso del trabajo de campo? Respondemos a esta pregunta mediante el análisis de dos indicadores clave: los puntajes de QuAC y las cancelaciones totales con respecto a la tendencia de cada entrevistador en el proyecto¹³. Debido a que nuevos entrevistadores comienzan a trabajar en el proyecto una vez que el trabajo de campo ya está en marcha¹⁴, evaluamos los puntajes de QuAC y cancelaciones en función del tiempo que un entrevistador ha trabajado en el campo¹⁵. Si bien las tendencias en las cancelaciones y puntajes de control de calidad varían un poco entre contextos e individuos, encontramos que, en promedio, tanto el porcentaje de entrevistas canceladas debido a la mala calidad como el puntaje promedio del QuAC disminuyen durante el trabajo de campo.

El Gráfico 2 muestra nuestros resultados respecto de los puntajes promedio del QuAC. En el panel izquierdo, la cantidad de días en campo de los entrevistadores se muestran en el eje X. Un valor de 1 indica el primer día de campo de un entrevistador, que podría ocurrir al comienzo del trabajo de campo nacional o después de que se hayan completado varios días o semanas de entrevistas. Los valores más altos indican días adicionales de experiencia entrevistando. Como se esperaba, los puntajes promedio de QuAC tienden a la baja a medida que avanza el trabajo de campo, aunque también hay picos notables en los puntajes de QuAC observados hacia el final del trabajo de campo.

El panel derecho del Gráfico 2 muestra la puntuación de QuAC predicha por un análisis de regresión lineal que incluye el número de días del entrevistador en el campo, un término cuadrático para capturar las no linealidades del entrevistador y efectos fijos por país. El análisis de regresión muestra, de manera similar, una tendencia a la baja en las puntuaciones de QuAC esperadas a lo largo del trabajo de campo, lo que indica una mejor calidad de la entrevista. Notamos que la gran mayoría de las entrevistas se llevan a cabo antes de la marca de los 100 días. Nuestras estimaciones de la calidad de la entrevista aumentan ligeramente después de la marca de los 100 días¹⁶. Creemos que esto refleja el hecho de que los equipos a veces se apresuran a realizar entrevistas al final del trabajo de campo con el fin de cumplir con los plazos establecidos. Esto puede generar un pico en las entrevistas de baja calidad (y cancelaciones) en los últimos días del trabajo de campo. Debido al número relativamente pequeño de entrevistas que tienen lugar en estas fechas tardías, vemos una incertidumbre sustancialmente mayor en torno al puntaje promedio de QuAC, sobre todo cuando la cantidad de días que un entrevistador ha pasado en el campo va más allá de aproximadamente 30 días.

Vemos una tendencia similar cuando consideramos el porcentaje de entrevistas que se cancelan cuando los entrevistadores trabajan en el proyecto. El Gráfico 3 muestra promedios descriptivos del porcentaje de entrevistas canceladas por día de trabajo (panel izquierdo) y un análisis de regresión idéntico al presentado en el Gráfico 2, utilizando el porcentaje de entrevistas canceladas como la variable dependiente (panel derecho).

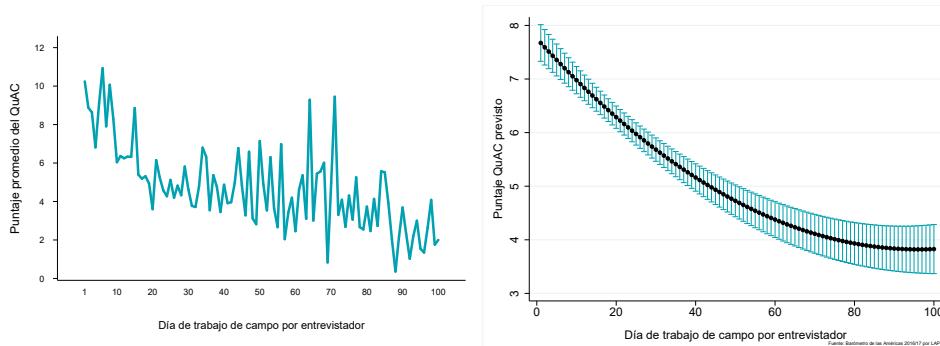


Gráfico 2: Puntuaciones promedio y puntuaciones esperadas del QuAC, por días de trabajo de campo por entrevistador

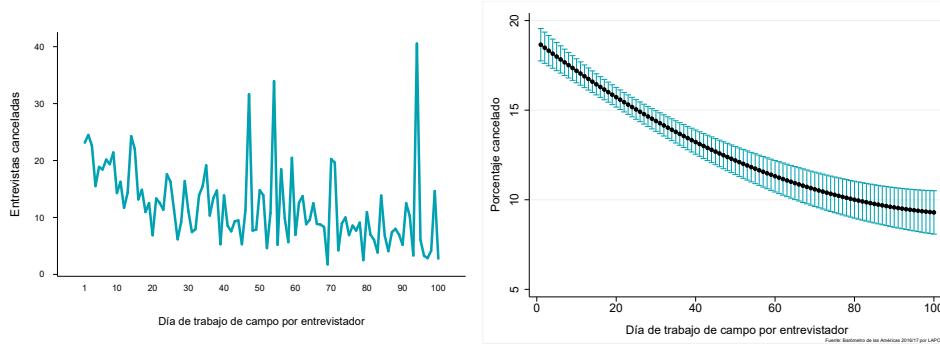


Gráfico 3: Cancelaciones promedio y cancelaciones previstas por días de trabajo de campo por entrevistador

Ambas cifras muestran una disminución en las entrevistas canceladas a medida que avanza el trabajo de campo, con niveles muy bajos de cancelaciones y picos repentinos en los últimos días del trabajo de campo. Como se mencionó anteriormente, creemos que esto es un reflejo de la prisa de los equipos por cumplir con los plazos del proyecto, lo que resulta en entrevistas de baja calidad que luego se cancelan. Al igual que el análisis descriptivo, el análisis de regresión muestra una tendencia decreciente en las cancelaciones, ya que un encuestador pasa más tiempo en el campo, con más incertidumbre sobre el nivel de cancelaciones hacia el final del trabajo de campo.

Es posible que la calidad de las entrevistas mejore en el transcurso del tra-

bajo de campo, independientemente de nuestros protocolos de control de calidad, a medida que las firmas locales separan a los entrevistadores de bajo desempeño y el resto de entrevistadores se familiarizan cada vez más con el cuestionario y, por lo tanto, pueden leer las preguntas con mayor fluidez. No podemos descartar esto a partir de estos datos; sin embargo, el aumento en los errores de los entrevistadores que observamos hacia el final del trabajo de campo sugiere que de hecho los entrevistadores pueden adherirse *menos estrictamente* a los protocolos estándar a medida que se familiarizan con el instrumento.

Fabricación de datos a lo largo del tiempo

Si bien la mayoría de las entrevistas canceladas se cancelan debido a lecturas pobres o errores accidentales, este no es siempre el caso. Ocasionalmente identificamos miembros de un “Equipo X”—aquellos que se niegan a seguir los protocolos después de que se les recuerda que lo hagan, o que hacen trampa de manera intencional y continua. Proporcionamos evidencia de este comportamiento a los equipos locales y solicitamos que estos entrevistadores sean separados del proyecto. Sin embargo, tales casos son raros. Solo alrededor del 1% de todas las entrevistas de estos diez países tuvieron errores consistentes con la fabricación de datos (7% del 14.3 % de las entrevistas que finalmente se cancelaron)¹⁷.

Además, los entrevistadores problemáticos generalmente se identifican en los primeros días del trabajo de campo. El Grafico 4 muestra el porcentaje promedio de todas las entrevistas con errores consistentes con fraude por día de trabajo de campo del entrevistador (panel izquierdo) y un análisis de regresión idéntico a los descritos anteriormente prediciendo el porcentaje de entrevistas con errores consistentes con fraude (panel derecho). Las tasas de entrevistas potencialmente fraudulentas son siempre bajas (menos del 2 % de todas las entrevistas). Sin embargo, la proporción de todas las entrevistas con errores consistentes con la fabricación de datos disminuye sustancialmente a medida que avanza el

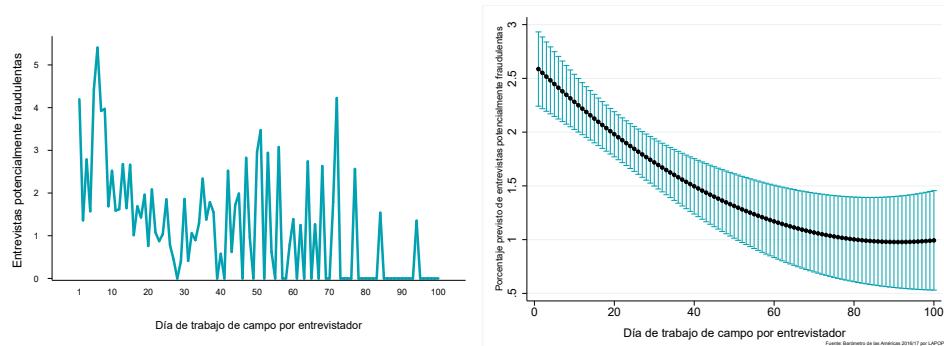


Gráfico 4: Porcentaje promedio y porcentaje previsto de entrevistas con errores consistentes con fraude, por día de trabajo de campo por entrevistador

trabajo de campo, de aproximadamente el 1.7% de todas las entrevistas en el primer día del trabajo de campo al 0.6 % de todas las entrevistas en el día 75.

No sólo son raras las entrevistas falsificadas, con algunas variaciones según el país y el entrevistador, sino que su prevalencia disminuye con el tiempo. Es decir, nuestra estrategia para erradicar entrevistas y entrevistadores de muy baja calidad resulta en una menor fabricación de datos a lo largo del tiempo. Sin embargo, un mensaje central de las cifras anteriores es que incluso cuando la fabricación está en su punto más alto, las entrevistas potencialmente fraudulentas representan una porción muy pequeña de todas las entrevistas realizadas como parte del proyecto del Barómetro de las Américas.

Conclusión

Al hacer un uso escrupuloso y sistemático de la grabación de audios, captura de imágenes y localización GPS, LAPOP puede evaluar y mejorar la calidad de las entrevistas y proporcionar retroalimentación oportuna y detallada a los entrevistadores y equipos de trabajo de campo en un corto tiempo. Continuaremos refinando estas herramientas y procesos en un esfuerzo por garantizar que los datos de LAPOP sean de la más alta

calidad posible.

Nuestros esfuerzos para identificar y codificar los diferentes tipos de errores que ocurren durante el proceso de entrevista también proporcionan una idea de la prevalencia de errores graves, como los consistentes con la fabricación de datos. Nos complace informar que tales errores representan una porción muy pequeña de todos los errores en nuestras entrevistas. La gran mayoría de los errores, como la lectura incorrecta de preguntas, son consistentes con entrevistadores descuidados u olvidadi-
zos, no con la falsificación de datos. Estos hallazgos son consistentes con análisis recientes (por ejemplo, Robbins 2016) que sugieren que los datos de opinión pública de contextos en desarrollo son generalmente con-
fiables. Además, estos son resultados alentadores para quienes trabajan con datos de opinión pública internacional.

Notas

1. Los protocolos de auditoría de LAPOP incluyen la auditoría humana descrita aquí, así como la implementación de guiones automatizados en nuestro software de recopilación de datos (en la ronda 2016/17, SurveyToGo) y R (consulte también el gráfico 4). En esta Nota Metodológica, “auditoría” se refiere al componente humano aquí detallado.
2. QuAC por sus siglas en inglés, es el “Capítulo de aseguramiento de la calidad”.
3. Todas las imágenes se capturaron utilizando una cámara frontal para garantizar que el anonimato del entrevistado no se comprometiera (es decir, la cámara solo capturó imágenes de los entrevistadores). Los participantes del estudio fueron informados antes de dar su consentimiento para ser entrevistados, que algunas de sus respuestas se grabarían con fines de control de calidad.
4. Antes de empezar el trabajo de campo, todos los entrevistadores del Barómetro de las Américas se someten a dos días de capacitación que cubre (entre otros temas) el proyecto, el cuestionario, las mejores prácticas y también incluye juegos de roles y entrevistas de práctica.
5. Además de esta revisión, analizamos periódicamente los “Indicadores Clave de Desempe-

ño”, KPI por sus siglas en inglés. Estos indicadores proporcionan información detallada sobre el inicio y finalización del trabajo de campo todos los días, entre otras métricas. Examinamos los registros de los dispositivos móviles para garantizar que los entrevistadores no hayan alterado la configuración del dispositivo para impedir, por ejemplo, la recopilación de las coordenadas GPS. También rastreamos la ruta que los entrevistadores recorrieron en el transcurso de cada día de trabajo de campo para evaluar si seguían correctamente las reglas para seleccionar las viviendas y los encuestados individuales. Estos protocolos también han mejorado nuestra capacidad de rastrear y verificar los intentos de contactos y rechazos. Cada uno de estos procesos se detalla en otra parte de la serie de Notas Metodológicas, por lo que no los analizamos aquí.

6. Actualizamos el QuAC en el transcurso de la ronda 2016/17. La segunda versión, evaluada aquí, se utilizó en Argentina, Bolivia, Brasil, Chile, Guatemala, Haití, Jamaica, México, Panamá, Perú y Uruguay.
7. Los equipo locales auditán el 100 % de las entrevistas durante todo del proyecto. Por ende, todas las entrevistas pasan por lo menos por una ronda de auditoría humana.
8. Usualmente se asigna un grupo de entrevistadores a un mismo auditor. De esta manera, el auditor individual se vuelve familiar con la voz del entrevistador, sus patrones al hablar, y nosotros creemos, se vuelve más hábil para identificar errores consistentes con fraude.
9. Revisamos periódicamente las entrevistas realizadas por miembros del “Equipo A” y, en el muy raro caso de que encontremos errores consistentes con la falsificación, llevamos a cabo una revisión exhaustiva de todos sus trabajos anteriores.
10. Si bien no podemos confirmar si los supervisores del equipo local transmiten nuestros comentarios a los propios encuestadores, sí encontramos que la calidad de las entrevistas realizadas por cada entrevistador tiende a mejorar con el tiempo, en consonancia con haber recibido la retroalimentación.
11. El “SIS” provee a los entrevistados la información sobre sus derechos como participante del estudio, incluyendo su derecho a no responder alguna pregunta y de terminar la entrevista en cualquier momento.
12. Dichos errores también son comunes entre las entrevistas que no fueron canceladas. 40.6 % de las entrevistas aceptadas tuvieron al menos un error de lectura, 16.4 % tuvieron errores relacionados con la hoja de información del estudio, 8.1% de las entrevistas aceptadas tuvieron problemas técnicos, y en el 7.7 % de las entrevistas aceptadas, los encuestadores omitieron secciones de preguntas.

13. Debido a inconsistencias en la información subyacente sobre el tiempo de la entrevista, Argentina es excluida de este análisis a lo largo del tiempo.
14. Nuevos entrevistadores se pueden añadir al proyecto durante el curso del trabajo de campo debido a rotación de entrevistadores. El momento de las entrevistas también puede ser escalonado en las regiones subnacionales, lo que lleva a fechas de inicio escalonadas para los entrevistadores.
15. Cuando no normalizamos los datos de esta manera, vemos más picos en ambas métricas, a medida que los nuevos entrevistadores comienzan el proyecto, cometen errores, se corrigen y mejoran con el tiempo.
16. El 90 % de todas las entrevistas en este subconjunto de países se completaron el día 102, y el 99 % de todas las entrevistas se realizaron antes de la marca de 150 días.
17. Los errores que podrían mostrar evidencia de fabricación de datos incluyen (pero no están limitados a): entrevistas que cambian de ubicación o se completan en una ubicación secundaria, entrevistas que tienen múltiples encuestados, entrevistas para las cuales la captura GPS ha sido desactivada manualmente y entrevistas que tienen lugar en entornos sospechosos (por ejemplo, un vehículo en movimiento o en un bar). En un pequeño número de instancias, capturamos a los entrevistadores que discutían sus planes sobre cómo eludir parte de la grabación de audio. Algunos de estos errores pueden tener causas legítimas. Por ejemplo, es posible que un encuestado pueda invitar a un entrevistador a un vehículo para completar una entrevista en un lugar privado, o que un entrevistador se dé cuenta de que olvidó ingresar su número de identificación exclusivo para completar una entrevista y lo hace en una ubicación secundaria después de completar la entrevista. Nos inclinamos hacia el lado de la precaución al cancelar dichas entrevistas. Por lo tanto, nuestra medida de entrevistas fraudulentas probablemente sea una sobreestimación.

Referencias

- Kuriakose, Noble, y Michael Robbins. 2016. “Don’t Get Duped: Fraud through Duplication in Public Opinion Surveys.” *Statistical Journal of the IAOS* 32.3 (2016): 283-291.
- Robbins, Michael. 2016. “Yes, You Can Trust International Surveys. Mostly.” *Monkey Cage*. http://wapo.st/21FBQb2?tid=ss_tw-bottom&utm_term=.0ad06aa88bc1.



La Dra. Mollie J. Cohen es investigadora postdoctoral de LAPOP. Ella diseñó la estrategia de auditoría humana para los estudios del Barómetro de las Américas para la ronda 2016/17.

Sebastián Larrea es asistente de investigación de LAPOP. Él coordina y supervisa el control de calidad y trabajo de campo para el Barómetro de las Américas.

Este reporte fue editado por el Dr. J. Daniel Montalvo, el Dr. Noam Lupu y la Dra. Elizabeth J. Zechmeister. La auditoría de este reporte fue hecha por el Dr. J. Daniel Montalvo. Este reporte fue traducido por Sebastián Larrea y Juan Camilo Plata. El formato, la producción, la revisión, los gráficos y la distribución del reporte fueron manejados por Rubí Arana, Emma Tatem y Zach Warner. Nuestros datos e informes están disponibles para su descarga gratuita en el sitio web del proyecto. Síganos en Twitter o Facebook para mantenerse en contacto.

Como miembro fundador de la iniciativa de transparencia de la Asociación Americana para la Investigación de la Opinión Pública (AAPOR), LAPOP se compromete a la divulgación rutinaria de nuestros procesos de recopilación de datos y presentación de informes. Se puede encontrar más información sobre los diseños de la muestra del Barómetro de las Américas en vanderbilt.edu/lapop/core-surveys.

Esta *Nota Metodológica* ha sido posible gracias al apoyo del pueblo de Estados Unidos de América a través de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) y Vanderbilt University. Las opiniones expresadas en esta *Nota Metodológica* corresponden a los autores y LAPOP y no reflejan necesariamente el punto de vista de USAID, del Gobierno de los Estados Unidos o cualquiera de las otras instituciones financieras. Las encuestas del Barómetro de las Américas de LAPOP son financiadas principalmente por USAID y Vanderbilt University. La ronda 2016/17 también tuvo el apoyo del BID, el PNUD, la Open Society Foundations y varios socios académicos e investigadores a lo largo de las Américas.