

Nuevas formas de muestreo para minorías y poblaciones ocultas: muestras por encuestado conducido en una población de inmigrantes sudamericanos*

New forms of sampling for minority and hidden populations: respondent samples conducted in a south american immigrant population

Recibido: octubre 2 de 2009 | Revisado: mayo 10 de 2010 | Aceptado: diciembre 17 de 2010

MANUEL CÁRDENAS CASTRO**
SUSAN YÁÑEZ YÁÑEZ***

Universidad Católica del Norte, Chile

RESUMEN

El Respondent-Driven Sampling (RDS) es un método de muestreo para poblaciones ocultas o de difícil acceso. Se trata de un procedimiento que permite realizar estimaciones de la representatividad de la muestra en aquellos grupos en que desconocemos el marco muestral. En este estudio, se presenta tanto la descripción teórica de dicha forma de muestreo como el reporte de su aplicación práctica en el caso de minorías étnicas (inmigrantes bolivianos, peruanos y colombianos) en el norte de Chile. El estudio comenzó con ocho semillas y se reclutó una muestra total de 109 personas provenientes de Colombia, Perú y Bolivia. La información resultante del procedimiento RDS informa de un patrón similar de reclutamiento entre hombres y mujeres, reflejando que los tamaños de las redes de ambos grupos no difieren significativamente.

Palabras clave autores

Respondent-Driven Sampling, poblaciones ocultas, minorías.

Palabras clave descriptores

Métodos de muestreo, poblaciones de difícil acceso, inmigración latinoamericana en Chile.

ABSTRACT

The Respondent-Driven Sampling (RDS) is a method of sampling for hidden or hard-to-reach populations. This is a procedure for estimating the representativeness of the sample in those groups that is unknown the sampling frame. This study presents both the theoretical description of this form of sampling as the report of its practical application in the case of ethnic minorities (immigrants Bolivians, Peruvians and Colombians) in northern Chile. The study started with 8 seeds and recruited a total sample of 109 people from Colombia, Peru and Bolivia. The information resulting from the procedure RDS reports a similar pattern of recruitment for men and women, reflecting the sizes of the networks of both groups did not differ significantly.

Key words authors

Respondent-Driven Sampling, hidden populations, minorities.

Key words plus

Sampling methods, Hard-to-reach sample, Latin American immigrants in Chile.

SICI: 1697-9267(201206)11:2<671:NFMMP0>2.0.TX;2-5

Para citar este artículo. Cárdenas, M. & Yáñez, S. (2012). Nuevas formas de muestreo para minorías y poblaciones ocultas: muestras por encuestado conducido en una población de inmigrantes sudamericanos. *Universitas Psychologica*, 11(2), 571-578.

* Este estudio ha sido financiado por el Fondo Nacional de Desarrollo Científico y Tecnológico (Proyecto FONDECYT 11070096)

** Profesor Asociado de la Escuela de Psicología de la Universidad Católica del Norte, Chile. E-mail: jocard@ucn.cl. ResearcherID: Cárdenas, M. F.3457-2012.

*** Investigadora en el Centro de Atención Psicosocial a Inmigrantes de la Escuela de Psicología de la Universidad Católica del Norte, Chile. E-mail: syanez@ucn.cl

Introducción

La recolección de datos sobre una muestra que puedan ser considerados representativos de una determinada población, es una de las ventajas que los tradicionales métodos de muestreo y de las técnicas de estimación pueden proveer. Una muestra probabilística es aquella en la que cada persona de la población objetivo tiene una probabilidad conocida, diferente de cero, de ser seleccionada. Se trata de muestras que permiten realizar inferencias confiables sobre la población de la que son seleccionadas, todo esto dada su representatividad y el conocimiento de la magnitud del error. Los tradicionales métodos de muestreo requieren conocer el marco muestral, es decir, el número de sujetos que componen el universo dentro del cual tiene sentido seleccionar una muestra que sea representativa del mismo. En un número importante de grupos estas técnicas no son aplicables.

Uno de los problemas que atenta contra la validez de un estudio es el de trabajar con muestras no representativas, ya que impide generalizar los datos al conjunto de la población de la cual dicha muestra es parte. En un conjunto amplio de temas de investigación carecemos de un marco muestral de la población a la que nos dirigimos, por lo que desconocemos su real tamaño o magnitud, y nos es prácticamente imposible construir dicho marco para el muestreo, dada las particulares características de la población. Este es el caso de grupos con conductas de riesgo (drogadictos o personas infectadas con VIH/SIDA), con características estigmatizadas por la población (personas gay y lesbianas), de pequeño tamaño y con alta dispersión geográfica (minorías étnicas e inmigrantes), etc.

Una serie de novedosos métodos de muestreo ha intentado paliar los efectos que la no representatividad de la muestra pueden significar, y acercarse a muestreos lo más representativos posibles de las minorías con las que se trabaja. Entre estos nuevos métodos cabe destacar el Time-Space Sampling (Magnani, Sabin, Saidel & Heckathorn, 2005; Seaman, Lauby & Liebman, 2002; Stueve, O'Donnell, Duran, Sandoval & Blome, 2001), Venue-Based

Samplig (Muhib et al., 2001) y el Respondent-Driven Sampling (Heckathorn, 1997, 2002, 2007).

El Respondent-Driven Sampling (RDS) es una estrategia de muestreo especialmente diseñada para los casos en que el marco muestral o el tamaño del universo resultan desconocidos. Es por esto que resulta particularmente útil en el trabajo con poblaciones ocultas o de difícil acceso, que son aquellas en que, como ya se ha señalado, se desconoce el tamaño o cuyos márgenes son desconocidos para el investigador, y en que dada la condición minoritaria del grupo este puede manifestarse receloso para entregar información, dados sus comportamientos estigmatizados, mal evaluados socialmente o incluso considerados ilegales (Heckathorn, 1997, 2002).

El RDS ha sido diseñado para corregir algunas de las carencias metodológicas de los clásicos métodos de muestreo en cadena (bola de nieve o informantes clave), y lo hace por medio de la elaboración de un procedimiento matemático que permite estimar la representatividad de la muestra obtenida (por medio del análisis de redes), y sin desestimar que los propios miembros del grupo son los más aptos para reclutar a los otros miembros de la población. RDS permite calcular estimadores insesgados y el error estándar de los intervalos de confianza. Estos cálculos se realizan sobre la base de la información entregada por los participantes respecto a su relación con los reclutados, y en atención de la estimación que hace cada participante del tamaño de su red social (personas con las mismas características que él conoce) (Abdul-Quader, Heckathorn, Sabin & Saidel, 2006; Heckathorn, 2002). Es decir, con el RDS, durante el proceso de muestreo, se construye el marco muestral sobre la base de las estimaciones del tamaño de la red de cada participante.

De similar forma que otros métodos de muestreo en cadena, el RDS comienza por el reclutamiento de un número limitado de semillas y se expande en progresivas "olas" de reclutamiento de pares. Las semillas reclutan solo en la primera ola, los reclutados en ella toman la responsabilidad de traer nuevos participantes en una segunda ola y así sucesivamente. Todos los reclutados son personas con los cuales el reclutador tenía previamente una

relación, es decir, son parte de sus redes sociales. En RDS cada participante recluta una cuota fija de personas, y se incorpora un sistema de doble incentivo (usualmente se remunera a la persona cuando contesta el cuestionario o realiza la entrevista, y una vez que las personas reclutadas por él asisten para completar el cuestionario o para ser entrevistadas). En este sentido, las “semillas” funcionan como reclutadores temporales, ya que se les ha explicado el sentido del estudio y se les ha entregado un número limitado de cupones (entre tres y cinco), para invitar a personas que puedan ser incluidas en la investigación. Este límite impuesto al reclutamiento permite minimizar la influencia de las semillas en la composición final de la muestra (Magnani et al., 2005). Adicionalmente, se ha constatado que cadenas de reclutamiento más largas permiten incrementar la riqueza de la muestra.

La teoría asume que en un número suficientemente largo de olas, la composición de la muestra se estabiliza (cuando esto ocurre, la muestra es denominada de equilibrio), y que en muestras donde el número de vínculos son mayores se obtienen mejores resultados (Abdul-Quader et al., 2006). La estabilización de la muestra indica que se ha llegado a un estado en el cual los estimadores convergen alrededor de la muestra de composición estable y no cambia durante los subsecuentes ciclos de reclutamiento (se estima que luego de la quinta ola de reclutamiento la muestra debería alcanzar dicho estado de equilibrio). Este criterio de saturación, por estabilización de la presencia porcentual de una serie de categorías grupales significativas para el estudio (sexo, grupos de edad, etc.), es el que nos indicará cuándo debemos detener el proceso de reclutamiento.

En términos generales, se puede afirmar que para la aplicación del RDS se deben cumplir ciertos supuestos básicos sobre la población objetivo (Heckathorn & Jeffri, 2005; Mantecón, Juan, Calafat, Becoña & Román, 2008), a saber: 1) los informantes (semillas y participantes posteriores) deben reconocerse unos a otros como parte de la población objetivo, de modo que posean los vínculos necesarios para reclutar a los miembros de la minoría; 2) deben poseer redes sociales suficiente-

mente densas, con el fin de garantizar cierta “profundidad sociométrica” y 3) la población no debe estar excesivamente segmentada, de modo que las oleadas no terminen por sobreincluir a miembros de un determinado subgrupo.

La lógica del RDS implica que se hagan estimaciones sobre la población, a partir de los datos correspondientes al tamaño de la red social de la muestra estudiada. Para ello, resulta necesario que cada participante del estudio pueda estimar el tamaño o número de relaciones que mantiene al interior de la población de interés. Es a través de dicha información que se hace posible la estimación de la prevalencia de algún rasgo específico. Justamente para ello que se deben calcular los intervalos de confianza sobre las estimaciones de dicha prevalencia (Salganik, 2006), la cual provee el rango dentro del cual el investigador espera encontrar el verdadero valor con algún nivel de confianza, por medio de la generación de muestras réplica creadas por un proceso de auto-reposición o *bootstrap* (Salganik, 2006; Salganik & Heckathorn, 2004). Este procedimiento consiste en la construcción de numerosas muestras de réplica a partir de la muestra observada, las cuales son utilizadas para producir un conjunto de estimaciones replicadas. Es justamente la variación de estas estimaciones la que permite construir el intervalo de confianza en torno a la estimación original. Para realizar este procedimiento, actualmente contamos con el programa RDSAT (Respondent-Driven Sampling Analysis Tool, el cual puede ser descargado gratuitamente en la página del autor).

El objetivo del trabajo es realizar un procedimiento RDS aplicado al caso de muestras de personas inmigrantes provenientes de países de Suramérica (Colombia, Perú y Bolivia), y en las cuales carecemos de un marco muestral completo y fiable. Para los fines de la presente investigación, se compararon hombres y mujeres inmigrantes, ya que en el estudio del que es parte este procedimiento de muestreo se trabajó con dichos grupos, contrastándolos en una serie de variables asociadas al bienestar y la calidad de vida. La idea central es conocer si el tamaño de la muestra obtenido, autoriza a los autores para realizar afirmaciones sobre la población

de inmigrantes, en lo referido a las diferencias entre hombres y mujeres.

Resultados

Procedimiento y participantes

La población objetivo estuvo conformada por inmigrantes provenientes de Perú, Bolivia y Colombia, con edades comprendidas entre los 18 y los 65 años, y que llevasen residiendo en este país a lo menos 6 meses. Esta población se considera oculta, dado el carácter estigmatizado de su procedencia y los altos niveles de prejuicio existentes en la zona sobre estos tres grupos. Adicionalmente, se trata de una población cuyo tamaño es difícil de determinar, dada la situación de irregularidad de buena parte de la misma.

A diferencia de los tradicionales procedimientos de RDS reportados, en este estudio las personas incorporadas no recibieron pago por su participación. Los datos fueron analizados mediante el programa RDSAT versión 6.0 y se siguieron las normas éticas de estudios con personas de la Comisión

Nacional de Investigación, Ciencia y Tecnología (CONICYT).

El trabajo de campo dio lugar a ocho olas (o cadenas) de reclutamiento, comenzando por ocho “semillas” y entrevistándose un total de 109 participantes (incluidas las semillas). En la Figura 1, se puede observar la cantidad de reclutados por ola. En la primera ola se entrevistaron ocho personas, posteriormente, entre segunda y cuarta, se produjo el mayor ingreso de participantes a la muestra. Luego de la quinta ola, comienza a disminuir la llegada de reclutados.

En la Figura 2, se puede observar la dinámica de reclutamiento de ambos géneros. El mayor porcentaje obtenido en el reclutamiento de mujeres ocurre en la primera y en la sexta ola, este corresponde a un 62.5 %. En el caso de los hombres, el porcentaje más alto de ingreso ocurre en la segunda ola con un 68.4 %. La última ola de encadenamiento ofrece un 57.1 % de mujeres entrevistadas y un 42.8 % de hombres encuestados. Esto resulta apropiado para el objetivo de conformación del marco muestral, ya que inicialmente cinco de las semillas eran mujeres y tres eran hombres, por lo tanto, se puede inferir que aun cuando la cantidad de semillas no era

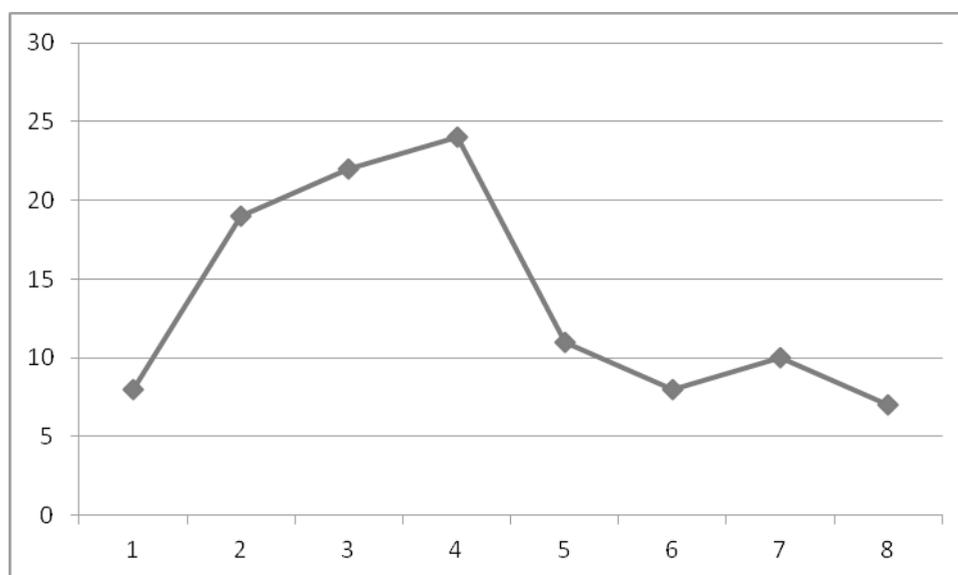


Figura 1. Entrevistados en cada una de las olas de reclutamiento.
Fuente: elaboración propia.

equitativa, se produce de todas formas la inclusión de ambos géneros.

Características demográficas

La muestra quedó compuesta por 109 personas, 59 hombres (54.1 %) y 50 mujeres (45.9 %), cuyas edades fluctuaron entre los 18 y los 65 años ($M = 35.01$; $DE = 11.35$). Del total de participantes, 33 procedían de Bolivia (30.3 %), 43 de Perú (39.4 %) y 33 de Colombia (30.3 %). Todos ellos tenían a los menos seis meses de permanencia en Chile ($M = 6.36$; $DE = 10.91$) y residían en las ciudades de Calama (46.8 %) y Antofagasta (53.2 %).

Del total de encuestados, el 33.9 % se encontraba casado al momento de la recogida de los datos y el 71 % refirió tener pareja en Chile (primando los vínculos con personas de su mismo país). La mayor parte declara tener hijos (65.1 %) y vivir con a lo menos uno de ellos en la ciudad (60.6 %). Por otra parte, el 85.3 % de los participantes se considera una persona religiosa. Respecto del nivel educativo de la muestra, se puede señalar que un 74.3 % ha

terminado sus estudios secundarios o tiene estudios técnicos completos (de este porcentaje, un 6.4 % cuenta con educación universitaria completa). En el ámbito laboral se puede observar que sobre el 52 % cuenta con algún nivel de calificación técnico o administrativo en su país de origen, sin embargo, en Chile, el mayor sector donde se desempeñan (45.9 %) corresponde a trabajos menores o bien oficios no calificados (construcción, servicio doméstico y comercio).

En relación con la situación jurídica de los participantes, un 38.5 % cuenta con visa temporal y solo el 27.5 %, con visa definitiva. Además, un 22 % posee visa sujeta a contrato (cerca de un 10 % se encuentra en condición de irregularidad). Finalmente, los motivos más reportados como causa de la migración son los referidos a mayores posibilidades laborales y mejora en las condiciones de vida personal y familiar (77.1 %). Otras motivaciones para la migración son la reunificación familiar (6.4 %), conocer otras culturas (4.6 %), persecución política (3.7 %), estudios (3.7 %) y amor (2.8 %).

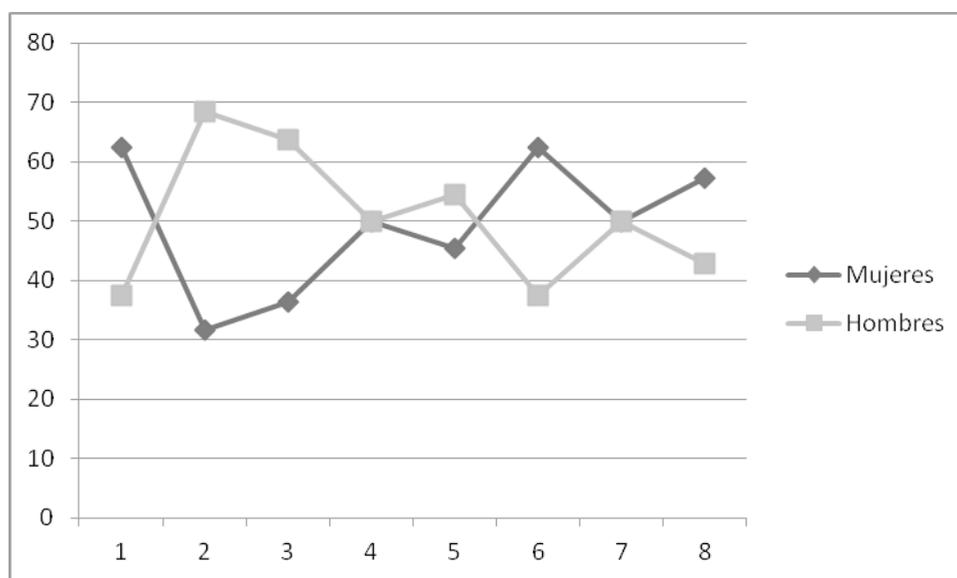


Figura 2. Evolución del porcentaje de entrevistados por género en cada una de las olas.
Fuente: elaboración propia.

Resultados del proceso de reclutamiento y tamaño de las redes

En las Tabla 1 se describen los resultados del reclutamiento de la muestra utilizada (con exclusión de las semillas).

TABLA 1
Recuento por género

Género reclutador	Género reclutado		
	Femenino	Masculino	Total
Femenino			
Recuento de reclutados	21	26	47
Proporción muestral (S)	0.447	0.553	1
Masculino			
Recuento de reclutados	24	30	54
Proporción muestral (S)	0.444	0.556	1
Distribución total de los reclutados	0.45	0.56	101
Distribución muestral (SD)	0.446	0.554	1
Muestra en equilibrio (E)	0.445	0.555	
Diferencia media entre SD y E		1.4%	
Tamaño medio de la red ajustado (N)	25.0	31.228	
Estimación de la proporción poblacional (P)	0.501	0.499	
Homofilia (H)	-0.108	0.113	

Fuente: elaboración propia.

Los resultados muestran que 21 mujeres han sido reclutadas por otra mujer y que 26 hombres también fueron reclutados por una mujer. Por lo tanto, se puede decir que el 44.7 % de las mujeres entrevistadas fueron seleccionadas por alguien de su mismo género y el 55.3 % de los hombres fue reclutado por una mujer. Por otra parte, 24 mujeres de la muestra fueron reclutadas por un hombre y 30 entrevistados fueron seleccionados por un hombre. Es así que el 44.4 % de las mujeres fue seleccionada por un

hombre y el 55.6 % de los hombres fue seleccionado por alguien de su mismo género. Estos porcentajes correspondiente a la proporción muestral (S).

La distribución total de la muestra (excluyendo las semillas) expresa que finalmente se reclutaron 45 mujeres y 56 hombres, que corresponden al 44.6 % y al 55.5 % de la distribución muestral (SD).

Ahora bien, la obtención de la muestra en equilibrio es sumamente importante, ya que la diferencia media entre esta y la muestra real permite conocer la proximidad de la composición de ambas muestras. En este estudio, y considerando el análisis por sexo, se obtuvo un 1.4 %, por lo tanto, existiría una aproximación muy cercana entre la muestra teórica en equilibrio y la muestra real. Esta diferencia es bastante buena, ya que su valor es menor al 2 %, el cual ha sido fijado como criterio para la toma de decisión (Heckathorn, 2002, 2007).

El tamaño medio de la red ajustado se obtiene al registrar el número de personas de la población objetivo que conoce el entrevistado (tamaño de la red social de personas de la misma característica de la persona que responde el cuestionario, para este caso, inmigrantes); en este estudio, se obtuvo un mínimo de 7 y un máximo de 389 personas, por esto, se determinó un límite que permitiera ajustar la red. Así, el tamaño medio de la red ajustado para las mujeres es de 25 personas y el de los hombres es de 31.2 personas. Se entiende que estas personas a las que hace referencia la red, son hombres o mujeres inmigrantes residentes en la II región, mayores de 18 años y que tienen un tiempo de permanencia mínima de seis meses en la zona.

Por otra parte, la *homofilia* corresponde al grado de afiliación o tendencia de reclutamiento dentro del mismo grupo de pertenencia. Según los resultados obtenidos, la incorporación que hacen las mujeres de nuevos informantes determina la selección solo en un 10.8 %, o dicho de otra forma, el 89.2 % de los nuevos participantes se produce de manera aleatoria. En el caso de los hombres el sesgo por homofilia interviene en un 11.3 %, por lo tanto, el 88.7 % de los nuevos participantes son reclutados de manera aleatoria.

Finalmente, en la Tabla 2 podemos observar el intervalo de confianza para la estimación calculado

TABLA 2
Intervalo de confianza

	Estimación de la proporción poblacional, P	Límite inferior	Límite superior
Masculino	0.499	0.41	0.605
Femenino	0.501	0.395	0.59

Fuente: elaboración propia.

por el programa RDSAT, con un nivel de confianza del 95 % ($\alpha = 0.05$), y que corresponde a los valores especificados en dicha tabla. Es importante destacar que en este estudio se trabajó sobre la base de la producción de 2.500 muestras réplica que brinda el RDSAT.

En conclusión, podemos decir que este procedimiento resulta beneficioso, cuando se desea obtener resultados aceptables, a pesar de que la muestra obtenida no sea probabilística (por los inconvenientes antes mencionados, es decir, por tratarse de una población oculta o de difícil acceso). Es importante destacar que las fortalezas de este tipo de estrategia muestral, se relacionan con la posibilidad de realizar análisis más exhaustivos sobre las redes (densidad y extensión), así como mejorar los muestreos por conveniencia, en cadena o teóricos, ya que al limitar el número de nuevos reclutados por participantes y registrar la red de personas con similares características que cada participante conoce, favorece realizar mejores estimaciones, además de optimizar la sistematización y seguimiento del proceso de reclutamiento por esta vía.

Discusión

El RDS ha mostrado ser una poderosa herramienta de muestreo en poblaciones ocultas o de difícil acceso. Es importante destacar que las fortalezas de este tipo de estrategia muestral se relacionan con la posibilidad de realizar análisis más exhaustivos sobre la densidad de las redes (densidad y extensión), permitiendo estimar el tamaño de la población, así como mejorar los muestreos por conveniencia y en cadena, ya que al limitar el número de nuevos re-

clutados por participantes se disminuye el riesgo de sobrerrepresentación de las redes. Por otra parte, el registro de la red de personas con similares características que cada participante conoce, favorece realizar mejores estimaciones, además de optimizar la sistematización y seguimiento del proceso de reclutamiento por esta vía.

En conclusión, la principal utilidad y fortaleza de este tipo de muestreo es que permite el acceso a poblaciones marginadas, ya sea de manera real o simbólica, facilitando que, más allá de la obtención de representatividad de una “pseudopoblación”, los grupos ocultos den a conocer sus experiencias y se pueda inferirla como común a todo el grupo al cual pertenecen. Saber si se pueden hacer afirmaciones sobre una población a partir de la muestra seleccionada, es fundamental y RDS permite dar un paso importante en este sentido.

Para efectos del presente estudio, los resultados indican un patrón similar de reclutamiento por sexo. Es decir, tanto los hombres como las mujeres inmigrantes mantienen relaciones con otros hombres y mujeres de dicho grupo y son, por esto, capaces de reclutarlos como parte de una muestra. El tamaño medio de sus redes tampoco varía de forma significativa, lo que indica que ni al interior del grupo total ni al comparar a los hombres con las mujeres, se encuentra un aislamiento relativo, ya que poseen un número suficientemente amplio de relaciones. Sería relevante ver si este patrón se reproduce, al realizar la comparación por países o grupos de edad.

Los resultados obtenidos por medio de este procedimiento de muestreo aseguran que las conclusiones que se pueden obtener en este estudio, en lo relativo a la comparación de hombres y mujeres inmigrantes provenientes de Perú, Colombia y Bolivia, pueden ser generalizadas a la población de personas inmigrantes provenientes de dichos países.

Referencias

- Abdul-Quader, A., Heckathorn, D., Sabin, K. & Saidel, T. (2006). Implementation and analysis of Respondent Driven Sampling: Lesson learned from the field. *Journal of Urban Health*, 83(7), 1-5.

- Heckathorn, D. (1997). Respondent Driven Sampling: A new approach to the study of hidden populations. *Social Problems*, 44(2), 174-199.
- Heckathorn, D. (2002). Respondent Driven Sampling II: Deriving population estimates from chain-referral samples of hidden populations. *Social Problems*, 49(1), 11-34.
- Heckathorn, D. (2007). Extensions of Respondent-Driven Sampling: Analyzing continuous variables and controlling for differential recruitment. *Sociological Methodology*, 37(1), 151-208.
- Heckathorn, D. & Jeffri, J. (2005, 22 de junio). *Assessing the feasibility of Respondent-Driven Sampling: Aging artist in New York City* [Documento de trabajo]. Research Center of Arts and Culture. Consultado el 23 febrero de 2008, en http://www.tc.columbia.edu/centers/rcac/pdf/FeasRep_12.pdf
- Magnani, R., Sabin, K., Saidel, T. & Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden population for HIV surveillance. *AIDS*, 19(Suppl. 2), 67-72.
- Mantecón, A., Juan, M., Calafat, A., Becoña, E. & Román, E. (2008). Respondent-Driven Sampling: un nuevo método de muestreo para el estudio de poblaciones visibles y ocultas. *Adicciones*, 20(2), 161-169.
- Muhib, F., Lin, L., Stueve, A., Millar, R., Ford, W., Johnson, W., et al. (2001). A venue-based method for sampling hard-to-reach populations. *Public Health Reports*, 116(1), 216-222.
- Salganik, M. (2006). Variance estimation, design effects, and simple size calculation for Respondent-Driven Sampling. *Journal of Urban Health*, 83(7), 98-112.
- Salganik, M. & Heckathorn, D. (2004). Sampling and estimation in hidden populations using Respondent-Driven Sampling. *Sociological Methodology*, 34, 193-239.
- Semaan, S., Lauby, J. & Liebman, J. (2002). Streets and network sampling in evaluation studies of HIV risk-reduction interventions. *AIDS*, 4, 213-223.
- Stueve, A., O'Donnell, L., Duran, R., Sandoval, A. & Blome, J. (2001). Time-space sampling in minority communities: Results with young Latino men who have sex with men. *American Journal of Public Health*, 91(6), 922-926.