



UNIVERSIDAD DE LA REPÚBLICA
Facultad de Ciencias Económicas y de Administración
Instituto de Estadística

Aplicación de la estrategia de Muestreo *Respondent Driven Sampling* en el estudio de Población trans en Uruguay

Ana Coimbra, Juan José Goyeneche, Guillermo Zoppolo
Noviembre 2014

Documentos de Trabajo

Serie DT (14 / 01) - ISSN : 1688-6453

Aplicación de la estrategia de Muestreo *Respondent Driven Sampling* en el estudio de Población trans en Uruguay

Ana Coimbra ¹

Instituto de Estadística - Facultad de Ciencias Económicas y de Administración - UdelaR.

Juan José Goyeneche ²

Instituto de Estadística - Facultad de Ciencias Económicas y de Administración - UdelaR.

Guillermo Zoppolo ³

Instituto de Estadística - Facultad de Ciencias Económicas y de Administración - UdelaR.

RESUMEN

Cuando la población de interés es un grupo pequeño y objeto de estigmatización (enfermos de VIH, usuarios de drogas, trabajadores sexuales, etc.) es natural que no se cuente con un marco muestral y sea impracticable su construcción. En este caso, los métodos convencionales de muestreo no son aplicables y deben ensayarse otro tipo de soluciones. La estrategia denominada Respondent Driven Sampling (RDS, Heckathorn (1992)) ha tenido numerosas aplicaciones en los últimos años. Presentamos una descripción de dicha técnica, la aplicación a un caso concreto y una evaluación de los supuestos en que se funda.

Palabras claves: Poblaciones Ocultas, Referenciación en Cadena, Cadenas de Markov.

¹acoimbra@iesta.edu.uy

²jjgoye@iesta.edu.uy

³gzoppolo@iesta.edu.uy

1. Introducción

En ocasiones, cuando la población objeto de estudio es un grupo pequeño y objeto de estigmatización (enfermos de VIH, usuarios de drogas, etc.) es natural que no se cuente con un marco muestral y sea extremadamente difícil y costoso su construcción. En este caso, los métodos convencionales de muestreo de poblaciones finitas no son aplicables y deben ensayarse otro tipo de soluciones. En general, dichas soluciones, se apartan del muestreo probabilístico y suelen incorporar incentivos para la participación en la muestra. Para poblaciones de este tipo, denominadas “ocultas”, suele apelarse, entre otras⁴ a las técnicas basadas en muestras formadas por referencias en cadena (*chain referral samples*), que comparten como característica la manera de acceder a los individuos de la población, a través de sus propios integrantes, partiendo de un conjunto arbitrario de ellos.

En general, las muestras formadas por referencias en cadena no producen muestras aleatorias y por tanto no es posible la deducción de estimadores insesgados. En este sentido se dice que las muestras que permiten obtener estas técnicas introducen “sesgos”, ya que generan probabilidades de inclusión desconocidas y distintas para cada individuo en la población estudiada. Las diferencias en las probabilidades de inclusión se deben fundamentalmente a que

- estas dependen de la selección de la muestra de partida (que suele ser no aleatoria)
- el mecanismo de selección depende de las características (desconocidas) de la red social donde tiene lugar el mecanismo de reclutamiento. De esta manera, el diseño muestral queda fuera del control del investigador.

Desde las ideas originales del muestreo del tipo “bola de nieve” (*snowball sampling*, (Goodman, 1961)) hasta la actualidad se han intentado refinamientos de las técnicas de muestreo por referenciación en cadenas con el objetivo de derivar estimadores insesgados. La razón es la del artillero: muchas veces no hay otra forma de obtener información valedera que permita investigar a este tipo de poblaciones. Uno de estos intentos, que parece haberse ido consolidando, tanto desde el punto de vista teórico, como del de sus aplicaciones prácticas, es el denominado *Respondent Driven Sampling* (RDS) propuesto originalmente en (Heckathorn, 1997) y ampliado en (Salganik and Heckathorn, 2004) y (Volz and Heckathorn, 2008).

⁴Muestras basadas en informantes Calificados (*Key informant sampling*) y las muestras orientadas (*Targeted Sampling*), son ejemplos de mecanismos que permiten observar a un subconjunto de individuos de una población dada pero sin aspirar a realizar inferencias basadas modelos probabilísticos

2. Antecedentes

A nivel internacional son muy numerosos y variados los casos en donde el RDS ha sido aplicado. Por ejemplo, la página web <http://www.ncbi.nlm.nih.gov/pubmed> del National Center for Biotechnology Information de la National Library of Medicine registra 562 entradas correspondientes a la búsqueda de publicaciones con la clave “respondent driven sampling”, y esto solamente refiere a asuntos vinculados a la medicina.

A nivel nacional solamente se conoce una aplicación del RDS: *Seroprevalencia de VIH en usuarios de drogas derivadas de la hoja de coca en Montevideo y área Metropolitana* a cargo de la consultora Equipos MORI por encargo del Observatorio Uruguayo de Drogas, el Ministerio de Salud Pública y el Equipo Conjunto sobre SIDA de Naciones Unidas estudio que tiene como objetivos determinar la prevalencia de infección por VIH, describir las características socio demográficas, prácticas sexuales y de consumo de drogas, identificar factores de riesgo y caracterizar la modalidad de acceso a servicios entre usuarios de pasta base, crack y de otra denominación que tenga la cocaína fumable en residentes en Montevideo y área metropolitana. Hasta la fecha no se publicaron los resultados.

Sin embargo, existieron por los menos tres talleres sobre el problema general de estimación del total en poblaciones «ocultas» (en particular para las poblaciones de hombres que tienen sexo con hombres (HSH) en los dos primeros talleres y usuarios de drogas en el segundo, ambos en relación como grupos de alto riesgo en el contagio de VIH). En las tres ocasiones participaron investigadores del área de muestreo del Iesta. El primero⁵ de ellos fue orientado más hacia el método captura y recaptura y dio lugar al trabajo *Estimación de Tamaño de Población HSH Montevideo y Área Metropolitana - Uruguay*. Los otros dos talleres fueron específicos sobre RDS^{6 7} siendo éste último el que dió lugar al mencionado trabajo de Equipos MORI.

⁵Taller de entrenamiento destinado a crear capacidades técnicas locales en el uso de metodologías de estimación del tamaño poblacional en la población de hombres que tienen sexo con hombres (HSH): “Size Estimation Planning Workshop” dictado por la consultora Virginia Loo. Montevideo, 2009.

⁶Taller de desarrollo e implementación de protocolos para tomar una muestra de la población HSH en Uruguay, capacitación en RDS (Respondent Driven Sampling) realizada por los consultores Lúgía Kerr y Carl Kendall con apoyo de OPS, en el marco de los acuerdos realizados en la Comisión Intergubernamental de VIH/Sida del MERCOSUR. Montevideo 2009.

⁷Características y uso de la metodología Respondent Driven Sampling- RDS en estudio de seroprevalencia y conocimientos, actitudes y prácticas entre usuarios de drogas en el marco del proyecto Actualización de la información sobre epidemia de VIH entre consumidores de drogas de alto riesgo de zonas urbanas de Uruguay. La actividad fue organizada por la Junta Nacional de Drogas, el Ministerio de Salud Pública y el Equipo Conjunto sobre VIH sida de las Naciones Unidas en Uruguay. Montevideo 2011. El curso estuvo a cargo de Alfonso Silva Santisteban, basado en el trabajo: “Understanding the HIV/AIDS Epidemic in Transgender Women of Lima, Peru: Results from a Sero-Epidemiologic Study Using Respondent Driven Sampling” (Silva-Santisteban et al., 2012).

3. Descripción del RDS

Las características distintivas del RDS son:

1. Reconoce a la teoría de redes sociales como clave para entender el mecanismo de recolección de datos para un subconjunto de la población objetivo vía una cadena de referencias.
2. Se apoya en supuestos sobre el comportamiento de las redes que permiten utilizar procedimientos de muestreo sobre cadenas de Markov para obtener estimadores asintóticamente insesgados (Hastings, 1970).
3. Uso de estimadores y razones de estimadores apropiados bajo muestreo con remplazo en poblaciones finitas, los usualmente se denominados como estimadores del tipo Hansen-Hurwitz (Hansen and Hurwitz, 1943).
4. La participación en la muestra de los individuos de la población se basa en un sistema dual de incentivos, un incentivo por participar en las entrevistas (*primary reward*) y un incentivo por reclutar a nuevos participantes en el estudio (*secondary reward*). Este segundo incentivo, puede, incluso, ser dirigido para estimular una mayor participación de sub grupos mas difíciles de contactar.
5. El reclutamiento se realiza por los propios individuos de la población objetivo.

En definitiva, el RDS pretende ser un sistema riguroso de muestreo no probabilístico basado en cadenas referenciales de forma de permitir hacer inferencia estadística válida para una población objetivo, controlando las fuentes de sesgo usuales en este tipo de muestras mediante el apoyo en la teoría de muestreo sobre cadenas de Markov. De esta manera se producen muestras que son independientes de la muestra inicial y reducen los sesgos producidos por el sobremuestreo de la población más visible, permitiendo, bajo la adopción de supuestos sencillos (aunque no necesariamente apropiados), obtener estimadores asintóticamente insesgados. Una de sus principales ventajas es que no requiere de la existencia de un marco muestral.

Como se mencionó, el reclutamiento secuencial y referenciado de individuos se basa en un sistema dual de incentivos y los reclutadores son los propios integrantes de la población objetivo, de manera que los problemas de encubrimiento tienden a reducirse ya que es el propio individuo el que decide participar o no. La historia de reclutamientos queda registrada gracias a la entrega de cupones numerados. Se parte de un grupo inicial (o grupo

“semilla”) al cual se le entrega un número fijo de cupones para que reparta, según una regla bien especificada, entre sus contactos en la población. El procedimiento se repite para los contactos de las semillas y así sucesivamente hasta formar un número de olas apropiado. Adicionalmente el RDS requiere que se registre el grado de cada respondente, o sea, el número de potenciales contactos en la red del individuo.

Los supuestos necesarios para poder deducir estimadores asintóticamente insesgados pueden enumerarse como sigue

1. El grado (número de contactos directos de un individuo en la red a la que pertenece); los respondentes declaran con exactitud su grado en la red.
2. El reclutamiento es al azar; un individuo ya reclutado entrega sus cupones con probabilidades iguales dentro de su red de contactos, adicionalmente se supone (aunque no es estrictamente correcto) que el muestreo es con remplazo. En lo que sigue se supone que hay un solo cupón por individuo a efectos de simplificar el ejemplo.
3. Reciprocidad; las conexiones en la red son recíprocas, si el individuo a tiene entre sus pares a b entonces el individuo b tiene entre sus pares al individuo a , lo que implica que la cadena sea no direccionada. Este punto es la clave del modelo de reciprocidad en que se basa la construcción de los estimadores.
4. El proceso de reclutamiento es una cadena de Markov en las sucesivas olas y donde cada individuo es un punto del espacio de estados. En la práctica, cuando cada individuo reclutado recibe más de un cupón, es más fácil pensar que los estados de la cadena son una partición de la población en grupos, definidos según ciertas características demográficas. En cualquier caso, el espacio de estados es finito y la hipótesis de reciprocidad implica que la cadena sea irreducible, lo que produce que todos los individuos se comunican en un número finito de pasos. De esta manera hay una única distribución estacionaria que se alcanza de forma relativamente rápida (la convergencia se da en forma geométrica).

Si bien estos requerimientos pueden cumplirse aproximadamente para muestras grandes, son difíciles de sostener en muestras pequeñas o moderadas y/o en caso de pocas olas.

Bajo los supuestos anteriores, la distribución estacionaria de la cadena de Markov es la que resulta de asignar a cada individuo (o grupos de individuos) una probabilidad de selección proporcional a su grado en la red. Entonces, alcanza con monitorear la cadena hasta que se considere suficientemente cercana a la distribución estacionaria y, luego, ponderar los individuos por los inversos de sus probabilidades estimadas.

Si suponemos que x^* es el vector de probabilidades de la distribución estacionaria, debe cumplir que

$$1. x^* = \frac{d_i}{\sum_U \delta_i}$$

$$2. x^* = x^* \mathbf{P}$$

donde δ_i es el grado i -ésimo individuo reclutado (o del grupo de individuos), la suma en U refiere a la suma en la población de interés y \mathbf{P} es la matriz de transición de primer orden.

De esta manera, se controlan los sesgos típicos de las muestras por cadenas referenciales:

- Si las olas son suficientes como para estar cerca de la distribución estacionaria se elimina, por definición, el sesgo que pueda introducirse en la selección del grupo semilla
- por otra parte, se respetan las distintas probabilidades de participar en la muestra originadas en los distintos grados individuales en la red.

Adicionalmente, (Heckathorn, 2002) propone un suavizado de los datos de manera que la matriz de transiciones estimada se ajuste al supuesto de reciprocidad y exista una única solución numérica para la estimación de la distribución estacionaria. Dicho suavizado fue utilizado en los cálculos, en (Volz and Heckathorn, 2008) se argumenta como dicho ajuste mejora sensiblemente las estimaciones. Por último, (Heckathorn et al., 2002), proponen una forma nueva para la construcción de estimadores que permite estimar promedios poblacionales para variables tanto continuas como categóricas. (Dichos estimadores no fueron utilizados para el presente trabajo).

La manera de obtener estimadores de prevalencia poblacional, es decir para variables categóricas, se describe a continuación. La primer derivación, se debe a (Heckathorn, 2002) y puede presentarse sintéticamente, para el caso de dos grupos de reclutamiento -donde no existen problemas de sobredeterminación del sistema de ecuaciones-

4. Estimadores de Prevalencia

Se supone que la población se encuentra clasificada en dos grupos $X = A, B$.

Notación

N_A : número de elementos en la población del grupo A ,

n_A : número de unidades en la muestra del grupo A ,

T_{AB} : número de lazos o contactos de los individuos del grupo A con individuos del grupo B ,

C_{AB} : probabilidad de que un individuo del grupo A seleccione un individuo del grupo B , los elementos de la matriz de transición de primer orden, \mathbf{P} ,

R_A : número total lazos o contactos de los individuos del grupo A ,

δ_i : grado del individuo i en la red,

δ_X : grado medio de los los individuos del grupo X ,

r_{AB} : elementos de la matriz de reclutamiento estimada, o sea, número de individuos reclutados (en la muestra) del grupo B por individuos del grupo A .

El número total de lazos o contactos del grupo A es la suma de los grados de los individuos que pertenecen a dicho grupo, luego,

$$R_A = \sum_{i \in A} \delta_i = N_A \delta_A.$$

La probabilidad de que un individuo del grupo A elija a un individuo del grupo B se obtiene como,

$$C_{AB} = \frac{T_{AB}}{R_A}$$

análogamente para C_{AB} luego,

$$T_{AB} = N_A \delta_A C_{AB} \quad \text{y} \quad T_{BA} = N_B \delta_B C_{BA}.$$

Utilizando la hipótesis de reciprocidad, las últimas cantidades deben ser iguales y se tiene que dividiendo por el total de la población (desconocido), N ,

$$N_A \delta_A C_{AB} = N_B \delta_B C_{BA}.$$

Para obtener P_A y P_B se divide por el total de la población (desconocido), N ,

$$P_A \delta_A C_{AB} = P_B \delta_B C_{BA},$$

por último, con la restricción $\sum_X P_X = 1$ se tiene

$$P_A = \frac{\delta_B C_{BA}}{\delta_A C_{AB} + \delta_B C_{BA}}.$$

Similarmente se obtiene P_B .

4.1. Estimadores para \hat{C}_{AB} y \hat{C}_{BA}

El estimador de la probabilidad de que un individuo del grupo A (B) elija un individuo del grupo B (A) surge de la matriz de reclutamiento estimada, luego,

$$\hat{C}_{AB} = \frac{r_{AB}}{r_{AA} + r_{AB}}$$

y

$$\hat{C}_{BA} = \frac{r_{BA}}{r_{BA} + r_{BB}}$$

4.2. Estimadores para $\hat{\delta}_A$ y $\hat{\delta}_B$

El estimador del grado promedio en la red de un individuo es la razón de dos estimadores del tipo Hansen-Hurwitz y coincide con la media armónica al grado en la red cuando el muestreo se realiza de manera proporcional al tamaño (Volz y Heckathorn (2008)).

$$\hat{\delta}_A = \frac{n_A}{\sum_{i=1}^{n_A} \frac{1}{\delta_i}}$$

$$\hat{\delta}_B = \frac{n_B}{\sum_{i=1}^{n_B} \frac{1}{\delta_i}}.$$

4.3. Estimadores de prevalencia

De esta manera, los estimadores para las proporciones poblacionales de los grupos A y B son:

$$\hat{P}_A = \frac{\hat{\delta}_B \hat{C}_{BA}}{\hat{\delta}_A \hat{C}_{AB} + \hat{\delta}_B \hat{C}_{BA}}$$
$$\hat{P}_B = \frac{\hat{\delta}_A \hat{C}_{AB}}{\hat{\delta}_A \hat{C}_{AB} + \hat{\delta}_B \hat{C}_{BA}}$$

5. Estimación del total de la población

Como se dijo, el RDS permite estimadores asintóticamente para proporciones poblacionales, pero sus bases teóricas en el muestreo con remplazo no permiten estimar, N , el total de la población de interés. Para tal propósito se pensaba recurrir al método de captura y recaptura, donde el listado de individuos “capturados” iba a tomarse de los datos de profilaxis del Ministerio de Salud Pública. Dado el tamaño muestral final y las, para nada menores, dificultades del cotejo de dicha lista con la de la muestra observada, se optó descartar cualquier intento de estimación ya que sería demasiado inestable y de difícil interpretación o de difícil contextualización.

6. La implementación del RDS para la población trans en Montevideo

La aplicación de técnicas rápidas de evaluación de la población, por ejemplo, entrevistas con referentes de la población trans en Montevideo, se logró reclutar 10 semillas. En un principio se definió que se considerarían 3 semillas pertenecientes al grupo de trans que no se prostituyen, 4 del grupo de trans que ejercen la prostitución y otras 3 que también ejercen la prostitución pero no en la calle.

Con respecto a los incentivos, no debían ser muy altos, para que el incentivo monetario no fuera el único incentivo para participar en el estudio, ni muy bajos que no valiera la pena reclutar pares. Es así que luego de analizar los pro y los contra de los distintos montos, los mismos fueron fijados en las siguientes cantidades

- Incentivo primario: 100 pesos uruguayos. Se paga al finalizar la entrevista.

- Incentivo secundario: 70 pesos uruguayos por cada entrevista realizada de sus reclutados.

Con la participación y el reclutamiento de tres pares el monto total ascendía a 310 pesos uruguayos. El incentivo secundario se pagaba una vez que la trans reclutada participara en el estudio.

Para el diseño del cupón, se buscó que fuera atractivo, en base a las características a la población de interés. Se decidió, además, estipular una fecha de vencimiento de un mes, con la pretensión de agilizar la recolección de la muestra, pero no necesariamente fue así; luego de pasado algún tiempo se recibieron cupones vencidos.

Los cupones



Figura 1: Cupón estudio Población Trans en Uruguay

En cuanto al grado del individuo en la red, las preguntas elegidas para investigarlo fueron las siguientes

1. ¿Cuántas trans conoces, por la cara o por el nombre, que vivan en Montevideo y que te encontraste en las pasadas dos semanas?
2. ¿Cuántas trans conoces, por la cara o por el nombre, que vivan en Montevideo y que puedas encontrar en las próximas dos semanas?



Figura 2: Cupón estudio Población Trans en Uruguay

3. Entre las trans que mencionaste en las preguntas anteriores, ¿con cuántas podrías intentar comunicarte por teléfono en este momento?
4. Entre las trans que mencionaste en las preguntas anteriores, ¿cuántas estarían dispuestas a participar en este estudio si fueran invitadas por ti?

El trabajo de campo se llevó a cabo, en un principio, con un jefe de campo, 3 digitadores y 9 encuestadores. El mismo tuvo algunas dificultades, principalmente en cuanto a concretar las entrevistas. Muchas se agendaron fuera de la Facultad de Ciencias Sociales, lugar que en un principio se consideró adecuado aunque posteriormente, se llegó, incluso, a realizar algunas entrevistas en la calle (lugar donde ejercían la prostitución alguna de las trans). La dificultad de comunicación a la hora de explicar el funcionamiento de los cupones fue, a veces, una barrera para el fluido proceso de las entrevistas. En un momento se decidió que una vez que una trans repartiera sus cupones, el equipo de trabajo de campo recibiría los teléfonos de las posibles reclutadas para, de ser necesario, llamarlas y concretar las entrevistas.

En la experiencia en Perú, en 14 semanas reclutaron 450 trans, se plantaron 10 semillas y cada participante entregaba 3 cupones. En nuestro caso, luego de ocho meses de trabajo de campo se logró alcanzar una muestra de 132 trans.

7. Resultados

Con el uso del software RDSAT (Volz et al., 2012), software específico para el análisis de datos provenientes de este tipo de estudios, se obtienen los estimadores de prevalencia de los grupos de interés, el momento en el que se llega al equilibrio dependiendo de la variable seleccionada y los pesos poblacionales.

La representación de la red de la muestra obtenida

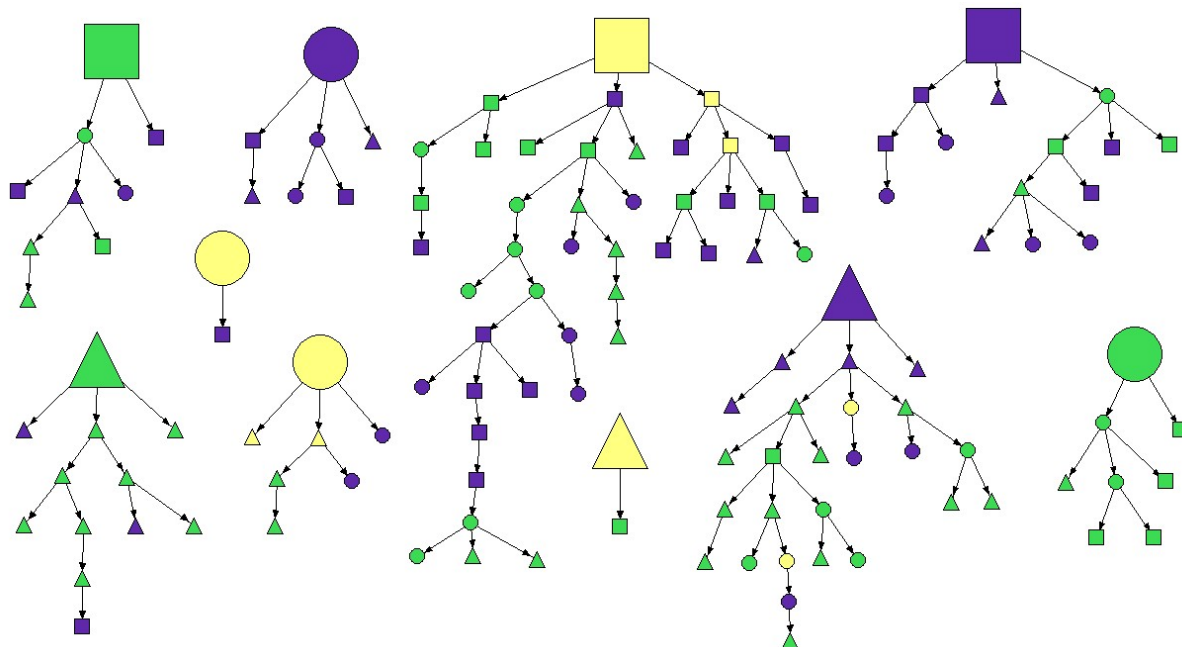


Figura 3: Muestra Generada

Para obtener ponderadores generales para la muestra observada se parte del promedio de los pesos individuales obtenidos para las variables respecto a prostitución y grupo de edad.⁸ luego los ponderadores son corregidos de manera de reproducir las variables mencionadas, simultáneamente en la muestra.

⁸El procedimiento de ajuste puede consultarse en (Deville et al., 1993). La implementación se realizó con el package survey (Lumley, 2004) del sistema R (R Core Team, 2014)

△	MENOR DE 32 AÑOS
○	ENTRE 32 Y 42 AÑOS
□	MAYOR DE 42 AÑOS
	NO PROSTITUCIÓN
	PROSTITUCIÓN CALLE
	PROSTITUCIÓN NO CALLE

Figura 4: Elementos que componen la red generada

No prostitución	Prostitución calle	Prostitución no calle
55	67	10

Cuadro 1: Cantidad de trans en la muestra por Clasificación

Menor de 32	Entre 32 y 42	Mayor a 42
49	44	39

Cuadro 2: Cantidad de trans en la muestra por Grupo de Edad

8. Conclusiones

En el caso en cuestión, el RDS es una de la pocas alternativas viables para acceder a datos sobre la población objetivo. Esta metodología e basa en supuestos razonablemente sostenibles cuando se obtienen muestras del orden de magnitud del que asegura resultados aceptables para un muestreo simple, aproximadamente unos 1000 casos, una presición de +/- 3% para una proporción poblacional (considerando el peor de los casos, cuando la población se divide en mitades), y un nivel de confianza del 95%, y no se descarta el empleo del RDS como técnica válida, aunque de dudosa aplicación al caso concreto, teniendo en cuenta el tamaño de muestra finalmente obtenido.

El magro resultado en términos de tamaño de muestra final seguramente tenga razones en, un posible sobredimensionamiento a priori del tamaño de la población, de un objetivo tal vez demasiado ambicioso, una etapa de preparación poco efectiva y un trabajo de campo con altibajos (sobre todo debido a la rotación -razonable debido a lo dilatado del tarbajo de campo- en supervisores y encuestadores. En este sentido, se asume una cuota importante de responsabilidad en tales fallas por parte de los investigadores del Iesta.

Se entiende que hay que tomar con mucha cautela las conclusiones que se puedan realizar con los datos obtenidos.

Por último, se valora muy positivamente el trabajo multidisciplinario que genera aspectos, para nada menores, como la construcción de un lenguaje compartido. También se considera positivo la obtención de una muestra de 132 trans, para la cual no es posible, por lo menos aún y en nuestra opinión, dimensionar su justa importancia

Referencias

- Deville, J., Sørnaldal, C., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88(1013-1020).
- Goodman, L. A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, 32(1961):148–70.
- Hansen, M. and Hurwitz, W. (1943). On the theory of sampling from finite population. *The Annals of Mathematical Statistics*, 14(4):333–362.
- Hastings, W. (1970). Monte carlo sampling methods using markov chain and their applications. *Biometrika*, 57:97–109.
- Heckathorn, D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44:174–179.
- Heckathorn, D. (2002). Respondent-driven sampling ii: Deriving valid population estimates from chain-referral samples of hidden population. *Social Problems*, 49:11–34.
- Heckathorn, D., Semaan, S., Broadhead, R., and Hughes, J. (2002). Extensions of respondent-driven sampling: A new approach to the study of injection drug users aged 18-25. *AIDS and Behavior*, 6(1):55–67.
- Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Salganik, M. J. and Heckathorn, D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, 34:193–239.
- Silva-Santisteban, A., Raymond, H., Salazar, X., Villayzan, J., Leon, S., McFarland, W., and Caceres, C. (2012). Understanding the hiv/aids epidemic in transgender women of lima, peru: Results from a sero-epidemiologic study using respondent driven sampling. volume 16, page 872.
- Volz, E. and Heckathorn, D. (2008). Probability based estimation theory for respondent driven sampling. *Journal of Official Statistics*, 24:79–97.
- Volz, E., Wejnert, C., Cameron, C., Spiller, M., Barash, V., Degani, I., and Heckathorn, D. (2012). *Respondent-Driven Sampling Analysis Tool (RDSAT) Version 7.1*. Cornell University, Ithaca, NY.