

EXTENDED ABSTRACT
Link-tracing Designs for the Study of Multi-Sited International Migration

M. Giovanna Merli
Sanford School of Public Policy and Duke Population Research Center
Duke University

Ted Mouw
Department of Sociology and Carolina Population Center
University of North Carolina, Chapel Hill

Claire Le Barbenchon
Sanford School of Public Policy
Duke University

Allison Stolte
Department of Sociology
Duke University

Extended Abstract

Large-scale, influential studies of international migration to the U.S. from Mexico and other Latin American countries (Massey and Zenteno 2000; Donato et al. 2010; Genoni et al. 2017) and to Europe from Africa (Beauchemin 2015), have illustrated the prominence of space, time and relationship between places of origin and destination for the study of migration. They have also highlighted the empirical challenges of accommodating these concepts within research designs and data collection on migration. First, international migrants are a rare population in the sense that they represent a small fraction of the population at destination (McKenzie and Mistaien 2009; Bilsborrow 2015). Second, migration involves places of origin and destination (Beauchemin 2014; Bilsborrow 2015, Massey et al. 1987; Parrado, McQuiston and Flippen. 2005). Third, migration is a network process and social networks factor into migration decisions (Côté et al. 2015; Baizan and Gonzalez-Ferrer 2016; Garip and Asad 2016; Toma 2016). Fourth, the migration process is transnational, that is migrants' situations can be fully understood through their participation in multiple places and transnational activities (Levit and Schiller 2004; Levit and Jaworsky 2007). These statements imply considerable challenges and potential limitations for the recruitment of useful samples that can adequately represent migrants in multiple sites, enable the analyses of the determinants and consequences of migration and focus on the connections between migrants' places of origin and destination.

First, obtaining samples of rare populations of migrants through conventional probability sampling designs for accurate representation requires costly, large-scale screening to generate samples of sufficient size for inference. Even when migrants are spatially clustered or disproportionately sampled from a sampling frame, obtaining population representative samples can be challenging because of the prevalence of illegal migration, as undocumented migrants may resist study participation for fear of repatriation.

Second, the analysis of the determinants and consequences of migration requires comparison groups selected at origin and destination in order to address the mechanisms underlying the selectivity of migration (Beauchemin 2014; Bilsborrow, Oberai et al., 1984). However, the issue of appropriate comparison groups (migrant and non-migrants in both origin and destination locations including before and after migration) continues to be salient in research design of migration studies (Bilsborrow 2015). This is an issue that requires large budgets, lengthy study duration and labor intensive follow-up, as highlighted by panel surveys that interview migrants and non-migrants before and after the migration and follow migrants overtime at destination (Rindfuss et al. 2007; Thomas et al. 2001, Genoni et al. 2017). Yet, even these longitudinal studies are often confronted by samples of migrants followed at destination that are too small or they cannot survey entire households who left their origin location unless they purposely recruit complementary samples of migrant households at destination.

Third, a focus on the social field where migration decisions and transnational activities are taking place or where immigrants undergo the process of integration and/or assimilation requires the challenging tasks of mapping networks that reach across multiple sites and cover the social field within which the transnational migration process unfolds (Lubbers, Verdery and Molina 2018).

In this paper, we describe the applications of link-tracing sampling designs, which can address some of the challenges to the collection of useful migrant samples. These designs collect referrals from respondents to track and interview future participants in the study. Sample recruitment starts with the identification of a limited number of initial respondents (“seeds”) known to the researcher, who are asked to provide minimally identifying information (e.g. initials of last name and full first name, last four digit of cell number, gender, or any other sufficiently detailed information) on friends and acquaintances who are members of their social network and who reside in the communities of destination or origin, depending on the purpose of the study (referred to in the social network literature as “alters”). Each seed is then asked to nominate a subset of alters who are invited to participate in the study, and to refer these nominated alters to the research team so that they can be contacted and recruited into the sample. This process continues through multiple waves of nominations, referrals and recruitment, until the desired sample size is reached. Typically, respondents are given space to nominate their alters with, what is referred in the literature as a “name generator,” a set of question(s) designed to elicit nominations about alters linked to respondents according to pre-defined relationship conditions. This is a common and well-studied method of eliciting socially relevant peers with reasonably high levels of validity despite some biases toward nominating peers who respondents interact with more frequently (Marsden 1993; Straights 2000; Marin 2004). Minimally identifying information collected as part of the network roster allows the combination of all nominated individuals into a single network and the identification of people who were nominated by more than one respondent.

In this paper, we discuss how variations on this sample recruitment approach was successively applied to studies fielded among different migrant populations and illustrate how this approach can address some of the challenges related to the recruitment of migrant samples: cost-

effectiveness, population representation, survey participation and incorporation of multiple sites into research designs of migration.

The first study we describe is a link-traced multi-site sample drawn from a well-connected community of Mexican migrants from Guanajuato, Mexico, to the states of North Carolina and Texas in the U.S. The sample consists of 600 members of this transnational community that spans these three regions, who were interviewed in person for the 2010 Network Survey of Immigration and Transnationalism (NSIT) (Mouw and Verdery 2012, Mouw et. al. 2014). We will rely on the data from this survey to show how the multi-site link tracing design has allowed to efficiently map the transnational field that connects places of origin and destination, examine its operation, and achieve high response rates in a population with a high prevalence of undocumented migrants.

The second study is a link-traced sample of Chinese immigrants to North Carolina consisting of 500 respondents interviewed using one of three modes of data collection (in person, phone and web) for the 2018 Chinese Immigrants in the Raleigh Durham Area (ChIRDU) study. The sample was collected using Network Sampling with Memory (NSM) techniques (Mouw and Verdery 2012; Merli et al. 2016). NSM is a link-tracing sampling design that collects detailed network data and uses these data to reveal the network at each sampling step. It uses a sampling algorithm to direct the sampling process to spread through the underlying network by placing higher sampling probability on nodes in the network who have been nominated less frequently by previously sampled respondents increasing the chance of discovering unknown sections of the network. The statistical properties of the NSM sampling algorithm have been tested with simulated sampling to show the accuracy, precision and efficiency of this approach in comparison to other better known link-tracing designs (e.g. Respondent Driven Sampling) and simple random sampling (Mouw and Verdery 2012). However, while simulated sampling is useful to test the statistical properties of the NSM sampling algorithm relative to simple random sampling, less is known about how this approach performs in the field and how it compares with conventional probability samples that claim to represent the population of interest. Here we will: (1) Show results of an evaluation of how the demographic characteristics of the ChIRDU sample compare with those of the American Community Survey and whether this comparison differs by mode; (2) Show results of an evaluation of spatial coverage of the ChIRDU survey. This evaluation relies on ChIRDU sample's geographic identifiers at the ZIP code level and on a adequately complete (Kennel and Li 2009) address-based list of the Chinese population of the Raleigh-Durham Area, obtained from an information reseller that provides addresses for direct marketing; (3) Use data collected in ChIRDU to demonstrate the feasibility of applying a name generator to collect multiple network rosters that expand the geographic coverage of ChIRDU respondents' networks and incorporate the multiple destination of other migrants in their networks.

Finally we will describe approaches to scale-up data collection and coverage of large geographic areas and multi-site populations.

References

- Baizan, P., and A. Gonzalez-Ferrer. 2016. What Drives Senegalese Migration to Europe? The Role of Economic Restructuring, Labor Demand, and the Multiplier Effect of Networks. *Demographic Research* 35(1):339–80.
- Beauchemin, C. 2014. A Manifesto for Quantitative Multi-sited Approaches to International Migration. *International Migration Review* 48(4):921–938.
- Beauchemin, C. 2015. Migration between Africa and Europe (MAFE): Looking beyond immigration to understand international migration. *Population-E* 70(1):007-012.
- Bilsborrow, R. 2015. Concepts, Definitions and Data Collection Approaches. In *International Handbook of Migration and Population Distribution*, M. White (ed). Berlin Heidelberg New York: Springer-Verlag.
- Côté, R. R., J. E. Jensen, L. M. Roth, and S. M. Way. 2015. The Effects of Gendered Social Capital on U.S. Migration: A Comparison of Four Latin American Countries. *Demography* 52(3):989–1015.
- Donato, K.M. J. Hiskey, J. Durand, and D.S. Massey. 2010. Eds. Continental Divides: International Migration in the Americas. *The Annals of the American Academy of Political and Social Science* 630:6-331.
- Garip, F., and A. L. Asad. 2016. “Network Effects in Mexico–U.S. Migration: Disentangling the Underlying Social Mechanisms.” *American Behavioral Scientist* 60(10):1168–93.
- Genoni, M.E., G. Farfan, R. Rubalcava, G. Teruel, D. Thomas, A. Velasquez. 2017. Mexicans in America. BREAD Working Papers 511.
- Kennel, T., and M. Li. 2009. Content and Coverage Quality of a Commercial Address List as a National Sampling Frame for Household Surveys. *Proceedings of the American Statistical Association, Section on Survey Research Methods* 2364–78.
- Levitt P. and N. Glick Schiller. 2004. Conceptualizing simultaneity: a transnational social field perspective on society. *International Migration Review* 38:1002–39.
- Levitt, P. and B. N. Jaworsky 2007 Transnational Migration Studies: Past Developments and Future Trends. *Annual Review of Sociology* 33:129-156.
- Lubbers, M., A. Verdery and J.L. Molina. 2018. Social Networks and Transnational Social Fields: A Review of Quantitative and Mixed-Methods Approaches. *International Migration Review*
- Marin, A. 2004. Are Respondents More Likely to List Alters with Certain Characteristics?: Implications for Name Generator Data. *Social Networks* 26/4: 289–307.
- Marsden, P. V. 1993. The Reliability of Network Density and Composition Measures. *Social Networks* 15: 399–421.

- Massey, D.S., R. Alarcón, J. Durand, and H. González. 1987. *Return to Aztlan: The Social Process of International Migration from Western Mexico*. Berkeley: University of California Press.
- Massey, D.S. and R. Zenteno. 2000. The validation of the ethno survey: The case of Mexico-US migration. *International Migration Review* 34: 766-793.
- McKenzie, D., & Mistian, J. 2009. Surveying migrant households: A comparison of census-based, snowball, and intercept surveys. *Journal of the Royal Statistical Society, Series A*, 172(2), 339–360.
- Merli, M. G., A. Verdery, T. Mouw, and J. Li. 2016. “Sampling Migrants from Their Social Networks: The Demography and Social Organization of Chinese Migrants in Dar Es Salaam, Tanzania.” *Migration Studies* 4(2):181–214.
- Mouw, T., S. Chavez, H. Edelblute, and A. Verdery. 2014. “Binational Social Networks and Assimilation: A Test of the Importance of Transnationalism.” *Social Problems* 61(3): 329–59.
- Mouw, T., and A. M. Verdery. 2012. “Network Sampling with Memory: A Proposal for More Efficient Sampling from Social Networks.” *Sociological Methodology* 42(1): 206–56.
- Parrado, E.A., C. McQuiston, and C.A. Flippen. 2005. Participatory Survey Research: Integrating community collaboration and quantitative methods for the study of gender and HIV risks among Hispanic migrants. *Sociological Methods and Research* 34: 204-239.
- Rindfuss, R. R., Kaneda, T., Chattopadhyay, A., & Sethaput, C. (2007). Panel studies and migration. *Social Science Research* 36, 374–403.
- Straights, B. 2000. Ego’s Important Discussants or Significant People: An Experiment in Varying the Wording of Personal Network Name Generators. *Social Networks* 22/2: 123–40.
- Thomas, D., E. Frankenberg and J. Smith. 2001. Lost but Not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey. *Journal of Human Resources* 36(3):556-592
- Toma, S. 2016. “The Role of Migrant Networks in the Labour Market Outcomes of Senegalese Men: How Destination Contexts Matter.” *Ethnic and Racial Studies* 39(4):593–613.