

## **Social Capital, Violence and Community Context**

### Research Methods and Data Collection

Communities within the city of Baton Rouge, Louisiana will be the focus of this study. Baton Rouge is a good study site because it has substantial heterogeneity in terms of race, social class, crime, and social capital across communities. All three members of the project team live and work in the city, and have been involved in conducting surveys in the city for many years. We intend to proxy communities within Baton Rouge using census tracts, a standard approach in the macro-social literature. There are approximately 90 census tracts within the broader community, providing a large enough sample to meet the axioms of the central limit theorem.

The data for this study will come from three main sources. Crime data will be secured from the local police department and geocoded to the local community level. Data from the US Census will be collated from publicly available files to provide socioeconomic and demographic data. The third source of data constitutes the bulk of this proposal, original survey data. For the survey data collection effort, we seek a large enough sample to have a reasonably valid portrait of attitudes and behaviors within communities. Our goal is to secure at least 50 full interviews within each census tract, and thus to have a completed survey sample size of between 4,500 and 5,000 interviews. This compares favorably to other major community studies, the most notable being the well known Project on Human Development in Chicago Neighborhoods.

In terms of the collection of the survey data, we have learned a great deal about data collection methods from our current work in post-Hurricane Katrina New Orleans, where we have conducted over 5,000 lengthy interviews, with good efficiency, among all population groups and neighborhoods. We had occasion to fundamentally re-think our approach to data collection, because Random Digit Dial (RDD) telephone sampling over land lines was so difficult in this setting. The following discussion lays out some of the lessons we have learned, and how these lessons inform the procedures we propose to use here.

*Description of the Survey Instrument.* Our survey attempts to locate individual respondents in their social networks and organizational environment. The survey contains several “hooks,” or merge variables to link individuals concretely to community- and higher level variables, so that we can build multi-level and geospatial models.

The survey instrument draws on previous research on social capital, social embeddedness, collective efficacy, confidence in public institutions and community organizations, nonprofit and faith-based organizations, and a variety of other fields. In particular, we use social capital indicators developed by Robert Putnam in his 2000 “Social Capital Community Benchmark Surveys” (Saguaro 2000). This was a multi-community survey, and Weil directed the Baton Rouge portion, with Blanchard’s collaboration, and has replicated the indexes in numerous subsequent Baton Rouge

and New Orleans surveys. Our survey instrument is strongly influenced by the work of Sidney Verba, with whom Weil studied, on civic engagement (Verba et al 1995), as was Putnam's before us. We also replicate several survey modules from Sampson (see for example Sampson et al.1999), on collective efficacy, intergenerational closure, and reciprocated exchange. Our base-line draft questionnaire is posted at <http://www.lsu.edu/fweil/LSUCrimeCommunitySurveyBatonRouge.pdf>. We will revise this draft on the basis of planned interviews with police and City-Parish officials, community leaders, and a limited amount of ethnographic interviewing.

*Need for a large sample.* In order to investigate the effect of group- or aggregated individual characteristics on a spatially distributed factor like violent crime, there must be sufficient individual-level measurements per spatial unit to compute spatial means. Otherwise, this causal direction cannot be investigated, but only the reverse: the effect of spatial or contextual factors on individuals. Thus, sample sizes of individual measurements must be quite large in order to compute means across a sufficiently large sample of districts. For instance, fifty interviews per district (to compute district means) across one hundred districts equals 5,000 individual interviews, which is the level of within unit density we look to obtain.

*Sampling Methods and Efficacy.* Random digit dial telephone surveys over land lines remain the norm in much survey research. Yet a standard RDD sample of thousands of respondents is problematical for several reasons: (1) It is very expensive to conduct thousands of telephone interviews, especially if the questionnaire is fairly long. (2) It has become increasingly difficult to draw a representative by RDD, especially because of the difficulty of reaching lower-income and young people, who are often not at home or do not have land lines.<sup>1</sup> Cell phone RDD samples have become more common – though by no means universal – but again, it is hard to keep people on a cell phone for even a fairly short questionnaire. Conducting an interview by cell phone costs the respondent both money (minutes) and cell-phone battery life. These hard to reach populations undoubtedly correlate strongly with the dependent variable, the rate of violent crime, and failure to reach them therefore seriously biases the sample.

Postal and self-administered paper-and-pencil sampling are less expensive, because interviewers do not have to be paid, and because paper questionnaires can be machine-read. Yet they share other weaknesses with RDD sampling for present purposes. Younger and lower-income populations are again likely to be hardest to reach by these methods. Residential mobility is high among both groups; and literacy is often a problem among lower-income people.

---

<sup>1</sup> Latest best estimates are that 17 percent of American households are cell-phone only, i.e., no land line. Among young people under 29 this reaches 31-36 percent cell-phone only; among people under the poverty line, 26 percent; and among renters, 34 percent. These percentages are rising rapidly (Blumberg et al. 2008).

Face-to-face surveys (door-to-door or at gatherings) are expensive<sup>2</sup> to conduct, but address many of the issues of representativeness, respondent cooperation, and literacy. Younger people are undoubtedly still more difficult to reach by this method, but probably no more difficult than with RDD sampling. Lower-income people are fairly easily reached by this method. Also, while face-to-face interviewing is difficult and expensive in large – national or most state – geographic settings, because of recruiting, training, and managing interviewers, plus their travel (unless a survey organization already has the interviewers in place), it is quite appropriate in smaller – local or community – geographical settings.

Finally, Internet surveys have the greatest potential problems of representativeness, for several reasons. First, as of the present writing, general probability samples are not yet commonly available for community-sized areas (city, county, zip code) from standard sampling companies.<sup>3</sup> Second, internet access is skewed to higher status and non-elderly people. However, internet samples can be highly effective for some populations, especially in cooperating communities with established, closed email lists. We have used this approach successfully in New Orleans with samples from the Jewish Federation (with comparisons to telephone samples from the same list) and some moderate-to-higher income neighborhoods. And data collection on internet is certainly the most cost-effective method. There are no expenses for each additional interview once the questionnaire has been posted; and unlimited questionnaires can be posted on commercial websites (e.g., SurveyMonkey) for a small annual fee, including unlimited responses and good sample and data handling.

The advantages and disadvantages of these data-collection and sampling methods can be summarized in Table 1, below. It seems apparent that there should be advantages to using a mixed method that maximizes representativeness and minimizes costs. In particular, the combination of internet and face-to-face data collection includes advantages with all population groups, averages out to moderate cost, and is appropriate for very large samples in smaller geographical settings. Our experience collecting data in post-Hurricane Katrina New Orleans suggests to us that this combination of methods will be effective in a setting like Baton Rouge, as we propose here.

---

<sup>2</sup> Computers are sometimes used to collect data in the field, to reduce costs, but from our experience in New Orleans, we do not believe they have a strong advantage. They are subject to malfunction, and they may not be appropriate in some settings. By contrast, paper-and-pencil questionnaires are very robust and can be machine read, thus possibly resulting in a lower overall cost.

<sup>3</sup> According to a discussion on AAPORNET (the official listserv of the American Association for Public Opinion Research) on August 7-9, 2009, any such sample might yield complete interviews in the single digits and would be nowhere as reliable as telephone RDD samples.

*Methods of Contacting Respondents and Eliciting Cooperation.* Implicit in our discussion of sampling methods is that success in contacting respondents and eliciting their cooperation also varies across methods and population groups. Two issues are especially central: issue-salience and the “chain of trust” between investigator and respondent. We have found in our post-Katrina research, both in New Orleans (the epicenter) and Baton Rouge (the nearest large “receiving” community of evacuees), that high issue salience strongly increases respondents’ willingness to cooperate and their patience with a long questionnaire. When the questionnaire is engagingly written and respondents feel the subject matter is important to them, especially locally, the likelihood of success is greatly increased. Issues of crime and safety are almost always considered to be salient; and we predict good respondent cooperation on this topic, even with a fairly long instrument.

**Table 1. Advantages and Disadvantages of Data-Collection/Sampling Methods**

Issue	Data Collection Method				
	RDD on Land Lines	RDD on Cell Phones	Postal and Self-Admin.	Face-to-Face: Door-to-Door or at Gatherings	Internet
Cost	<b>High</b>	<b>High</b>	<b>Low</b>	<b>High</b>	<b>Low</b>
Representativeness*					
Overall	Avg	Avg [ <b>Low</b> ]	<b>Low</b>	<b>High</b>	<b>Low</b>
Higher Income	Avg	Avg [ <b>Low</b> ]	Avg	<b>High</b>	<b>High</b>
Lower Income	<b>Low</b>	Avg [ <b>Low</b> ]	<b>Low</b>	<b>High</b>	<b>Low</b>
Younger People	<b>Low</b>	<b>High</b> [ <b>Low</b> ]	Avg	Avg	<b>High</b>
Older People	<b>High</b>	<b>Low</b>	Avg	<b>High</b>	<b>Low</b>
Appropriate for:					
Large Areas	Avg	Avg	Avg	<b>Poor</b>	Avg
Small Areas	Avg	Avg	<b>Good</b>	<b>Good</b>	Avg

\*Reachable and Cooperative. If they differ, Reachable is shown first, with Cooperative in [Brackets].

Secondly, the “chain of trust” between investigator and respondent is by no means guaranteed. These “chains of trust” can cross many different status lines, connecting investigators to respondents; and distrust at any link can render the enterprise unsuccessful. We have found in our work in post-Katrina New Orleans that, if the investigators take even a minimally active and constructive role in community affairs, it is quite feasible to find or build “chains of trust” from investigators, through community groups and leaders, to community members. Key here are honesty, transparency, and cooperative behavior on the part of the researcher. A number of population groups in New Orleans were unwilling to participate in our survey until they became persuaded that our intentions and efforts were genuinely helpful to the community. It would have been impossible to deceive these respondents, if our intentions and engagement had not in fact been helpful.

*Proposed Methods of Data Collection.* In light of these considerations, and in light of our experience in New Orleans, we propose to use a combination of internet, face-to-face, and self-administered data collection, verified by a smaller RDD sample, and with post-weighting of the sample as necessary.

We have worked closely with Baton Rouge Mayor Kip Holden for many years,<sup>4</sup> and he has agreed to help facilitate acquisition of crime data and of contacts to neighborhood associations. The City Government of Baton Rouge/Parish of East Baton Rouge has an Office of Neighborhoods (<http://brgov.com/dept/neighbors>); and a nonprofit organization also exists, the Federation of Greater Baton Rouge Civic Associations (<http://www.fgbrca.org/>). There are at least some 65-70 organized neighborhood associations – almost certainly more – and we propose to contact each of them and request that they send out an email to residents requesting participation in our survey on the internet. We will also send students door-to-door to leaflet this request. As noted above, we expect that this method will work well primarily in moderate to higher-income neighborhoods, and that issue salience and the “chain of trust” will work to our advantage. In lower-income neighborhoods, we would send student interviewers door-to-door to conduct the interviews face-to-face. We would also supplement these efforts with data collection at community events, through churches, and possibly through schools (students take questionnaires home to parents, who fill them out, and the students return them to the school, where we pick them up). Again, in lower-income neighborhoods, we expect that issue salience will also work to our advantage, and that by establishing good connections with community leaders, we can establish a good “chain of trust.” We have successfully collected thousands of interviews by these methods in Greater New Orleans, and have found them to be productive and cost-effective.

To summarize then, we plan to use RDD surveying to collect a sample of approximately 500 respondents against which to benchmark our data collected through other means. We then seek to collect between 4,000 and 4,500 through the other means described above, for a final sample of between 4,500 and 5,000 respondents.

*Representativeness and Bias in the Sampling.* We acknowledge that our proposed approach is not perfectly free from sampling bias, but (a) we do not know of another method that is freer of bias and still cost-effective, and (b) we take several measures to recognize and reduce sample bias. First, of course, if RDD sampling were sufficiently bias-free, we would use it, as is traditionally done. However, as we note above, we believe that it increasingly contains as much or more sampling bias than the methods we propose to use here; and of course, it is expensive. Full door-to-door

---

<sup>4</sup> Mayor Holden – and his predecessor before him – has come to Weil’s Sociology Methods class every year for nine years to receive a report on our class survey. We also reported the results of our Post-Katrina Baton Rouge survey to Mayor Holden, and he has often cited the results publicly.

sampling is perhaps freest of sampling bias, but it is extremely expensive. Second, our approach is not to focus so much on the representativeness of the full, aggregate sample, but rather, on the representativeness of each district or community sample. Our principle aim is to estimate district means, which we will then analyze against district crime rates, comparing one district to another. To be sure, we will develop post-sampling districts weights, as well as weights for the City/Parish as a whole, according to Census enumerations. (Census enumerations and estimates for Baton Rouge are considered to be reasonably reliable, in contrast to their estimates for post-Katrina New Orleans, which have been much disputed.) Third, by combining several different sampling methods in all districts, we provide internal checks to each method's biases. And finally, by focusing on variation among the important independent variables (esp. social capital) – and controlling for them in the analysis – we aim to reduce sampling bias on these variables, or to at least make them sufficiently explicit and known, that we are able to take them into account in our analysis. But to reiterate, and most importantly, if we believed there were a less potentially biased, yet cost-effective, approach, we would use it.

*Work cited*

Blumberg SJ, Luke JV. 2008. "Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2008." National Center for Health Statistics. Available from: <http://www.cdc.gov/nchs/nhis.htm>. December 17, 2008.