

SAMPLING and WEIGHTING for The China Survey

Pierre Landry (Yale University) and Mingming Shen (Research Center for Contemporary China, Peking University) were chiefly responsible for the sampling for The China Survey. The following description is provided by Landry in Pierre F. Landry and Daniela Stockmann, "Crisis Management in an Authoritarian Regime: Media Effects during the Sichuan Earthquake," paper prepared for delivery at the 2009 Annual Meeting of the American Political Science Association, Toronto, September 3-6, 2009, pp 6-7:

It is based on a stratified multi-stage probability sample of all Chinese adults. The sample uses county units as PSUs allocated in 16 strata. [Note: In China county-units refer to counties, autonomous counties, county-level cities, and urban districts.] We devised seven geographical regions, not including the four centrally administered municipalities that were treated as a separate stratum. [Note: The China Survey excludes the Special Administrative Regions of Hong Kong and Macao. The four CACs are Beijing, Tianjian, Shanghai and Chongqing] Each geographical region (including the central municipalities) was split between an urban and a rural stratum based on the share of rural residents in a given PSU. [Note: All units where this share exceeded .5 based on the 2005 population data were categorized as urban.] A total of 75 PSUs were selected according to the principle of probability proportional to measure of population size (PPS). SSUs consists of townships (or urban street committees), and two SSUs were selected in each county, also by PPS. The third stage consists of half-square minutes of latitude and longitude, drawn from a geographical grid of each SSU (township), excluding patently empty spaces identified *ex-ante* from high-resolution satellite imagery. The fourth stage consists of square seconds, the number of which is inversely proportional to the expected population density of each TSU. Trained surveyors equipped with GPS receivers were dispatched to systematically enumerate the final spatial units. [Note: For details of this process applied to survey in Beijing, see Landry and Shen (2005), "Reaching Migrants in Survey Research: the Use of the Global Positioning Systems to Reduce Coverage Bias in China," *Political Analysis*, 13: 1-22.] Finally, teams of trained interviewers were sent to interview all residents (one per household, randomly selected using the Kish grid method) of each spatial square second (SSS). This procedure yielded a national sample of 3989 respondents. Design-based estimates take PSUs, SSUs and TSUs into account, including the relevant finite population correction factor at each level. Given the spatially clustered nature of the final respondents, information beyond the TSU that may otherwise compromise the anonymity of the respondents is not included in the dataset. Thus, design-based estimates for China ignore clustering by SSS.

The data set includes a number of weighting variables. These are described in the RCCC's sampling report as follows:

Weighting and post stratification were done on three major variables, age, gender, and strata, as following.

First, calculate fpc of PSU, SSU, and TSU: fpc1, fpc2 and fpc3;

Second, calculate probability of every eligible adult to be selected into the sample. The variable name is "**wt_respo**".

Third, calculate the post-stratification weight in terms of strata, age(10 years), and gender based on the 2000 Census data. The variable name is "**wt_psf**".

We also calculate the post-stratification weight in terms of age(10 years), and gender. The variable name is "**wt_psbfc**".