Sample Analysis of a Population Survey for the 'Public Attitudes towards Nuclear Issues in Saskatchewan' Study

Jana Fried[1], Lauren Arnold[2], Scott Bell[3], Loleen Berdahl[5], Maureen Bourassa[6] and Ting Wei[7]

1 Jana Fried, Social Sciences Research Laboratories, University of Saskatchewan, Saskatoon, SK, jana.fried@usask.ca or Jana.Fried@coventry.ac.uk
2 Spatial Initiative, University of Saskatchewan, Saskatoon, SK, lauren.arnold@usask.ca
3 Geography, University of Saskatchewan, Saskatoon, SK, Scott.Bell@usask.ca
4 Political Studies, University of Saskatchewan, Saskatoon, SK, Loleen.Berdahl@usask.ca
5 Edwards School of Business, University of Saskatchewan, Saskatoon, SK, Bourassa@edwards.usask.ca
6 Department of Finance and Treasury Board, Government of Nova Scotia, Halifax, NS, tingweicn@gmail.com

Abstract

In order to investigate if there is a spatial pattern in response rates to a telephone-based provincial opinion survey, the data set of a population survey on public attitudes towards nuclear issues in Saskatchewan was used to calculate sampling errors for consolidated census subdivisions. An analysis of spatial autocorrelation among those sampling errors was conducted. The calculated Moran's I suggested only a slightly positive spatial autocorrelation, indicating that it is unlikely that there is a spatial pattern to non-response to the survey.

Background and Relevance

Surveys are the predominant tool used to measure public opinion. They allow the gathering of information about a sample of individuals in the hope of estimating characteristics for the whole population. The representativeness of the sample is influenced by both the sampling strategy (sampling error) as well as non-sampling errors (human error). One common strategy to reduce some of the sampling error is to increase sample size. However, increasing sample size does not solve the problem of non-response. Many subsets of a larger population are not evenly distributed geographically; when these sub-populations are important to the research/science question it is important to establish the extent to which they are under-represented in the survey sample (Bell, 2009).

According to Singh (2007), non-response is the "failure to obtain information for selected households or failure to interview eligible individuals" (Singh 2007, 111). Non-response would not be a problem if the likelihood of non-response is evenly distributed across different sub-groups of the total population. Sub-groups can be defined in myriad ways; some sub-groups will be essential to answering research questions, while others might only be important for ensuring overall representativeness. In this case we are interested in the geographic distribution;

secondarily, in our study area (Province of Saskatchewan) important sub-populations cluster in geographically definable parts of the province (farmers in rural areas and the south, aboriginals in segregated parts of cities and the north). Furthermore, non-response can introduce sample bias; over or under-sampling is possible in the total sample as well as in sub-population groups (age, gender, aboriginal people, geography, thematically interested versus not-interested population sub-groups) (Bell, Jones, & Wei, 2013; Bell & Wei, 2014).

A number of strategies exists that can reduce the effect of non-response. These include population comparison, comparison to external estimates, intensive post-sampling, and wave extrapolation (for more information, see e.g. Singh 2007). In the following section, we use the population (census data) comparison approach to test if our sample displays any spatial autocorrelation in the sampling error, i.e. if specific neighbouring areas within Saskatchewan were more likely to be under- or over-sampled. Spatial autocorrelation in the weights could potentially indicate that, for example, the topic of the survey had a spatially varying impact on the response rate to a particular survey.

Methods and Data

In the fall of 2013, the Social Sciences Research Laboratories at the University of Saskatchewan conducted a telephone survey of 1355 inhabitants of Saskatchewan. The survey provides insight into public attitudes towards nuclear issues (e.g. uranium mining, nuclear energy production, etc.) in Saskatchewan. The telephone survey was administered by trained student interviewers, using a random and representative sample of residents of the province. Individual interviews lasted approximately 15 minutes, with questions covering issues related to knowledge, perceptions of different nuclear sectors, future expectations, trust, and general worldview questions as well as some general demographic questions (Berdahl, Bell, Bourassa, & Jana, 2014; Bourassa, Bell, Berdahl, & Jana, 2014; Fried, Bell, Berdahl, & Bourassa, 2014a, 2014b). We also collected the full postal code for each individual respondent.

The overall response rate for the survey was 21%, i.e. a fifth of the total number of people called completed the survey. However, since we could not utilize postal code information for those people not agreeing to take part in the survey, we cannot directly calculate response rates for smaller spatial units. Hence, we calculated a weight as a measure of the sampling error and proxy for response rates based on the expected number and actual number of completed surveys for each geographic unit. While we report census consolidated subdivisions, we assessed a variety of units of analysis (e.g. census divisions, census subdivisions, forward sorting areas) in order to find a reasonable compromise between resolution (small enough units to support spatial analysis) and representativeness (large enough so that each unit would have participants). Consolidated census subdivisions do not fully complete these criteria but, given the survey size, present the best possible option to conduct this analysis.

Postal codes were used to assign individual cases to one of the 300 census consolidated subdivisions (CCS) of Saskatchewan (see Statistics Canada 2013). We then calculated a variable weight representing the sampling error for each CCS by using census data of the total population. This weight can be used for further statistical analysis of the survey data, e.g. to weight frequencies for other variables in our survey. The weight was calculated by dividing the expected number of completed surveys for each CCS by the actual number of completed surveys, with the expected number of surveys resulting from the percentage of Saskatchewan's population in each CCS (census data), multiplied by the total number of our completed surveys divided by 100.

In a second step, we used GeoDa (Anselin et al. 2006) to analyze the data for spatial autocorrelation in the sampling error (weight). We calculated both the global spatial autocorrelation index Moran's I and the local spatial autocorrelation index (LISA). Statistical tests are based on the notion that values of observations in the sample are independent of each other. The existence of spatial autocorrelation in our data would invalidate this assumption. Hence, in our case, spatial autocorrelation in our sampling error (weight) could hint at some spatially varying factors that influenced the sampling error differently for different geographical units.

Results

Results of our analysis are displayed in Figures 1 to 3, showing the calculated weights for each CCS; the Moran's I and regression line; and the local spatial autocorrelation index for each of the 300 CCS. Figure 1 shows a map of the 300 CCS of Saskatchewan with the calculated weights for each area ranging from 0 to 7 (results displayed in quintiles). The resulting values equal to 1 (n=61) indicate that the sample for the CCS is representative for the CCS' population as percentage of Saskatchewan (perfect relative sample size for that CCS), values smaller than 1 indicate an oversampling of the CCS, and values above 1 an under-sampling of the CCS, respectively. A weight of zero was assigned to those CCS for which we had zero completed surveys (see Figure 1, ArcGIS map). The large number of zero values (n=117; approximately a third of all CCS), representing no completed surveys or no expected surveys for the respective CCS, indicates that our overall sample size might not be sufficient to allow for a statistical analysis of our data at the CCS unit level.

Spatial patterns were also examined visually. When looking at just the map in Figure 1, no specific spatial pattern seems to appear, indicating spatial autocorrelation might not exist. To test this, we can refer to Moran's I for the total data set (calculated using queens contiguity). Using queens contiguity ensures that all contiguous neighbours are included in calculating the weighted average necessary for calculating spatial autocorrelation in the target variable (Anselin, Syabri, & Kho, 2006; Shah, Aspen, & Bell, 2014; Vivier et al., 2011). The value of Moran's I expresses the correlation between our variable weight and the spatial lag for the same variable. Moran's I is 0.17553 indicating a very slight positive spatial correlation for the total sample (see Figure 2).

However, the global index Moran's I does not communicate what regions are contributing to the global pattern (i.e. clusters of high or low values). Here, a calculation of LISA can be used to display local spatial autocorrelation patterns. While the majority (n=256) of CCS did not display significant LISA values, it is interesting to see that there are both clusters of under- (HH) and oversampled (LL) CCS (see Figure 3). In this context it is also noteworthy that nearly double the number of CCSs were oversampled (n=79) compared to those that were undersampled (n=43) (this ignores those CCS for which the weight was exactly 0 or 1).

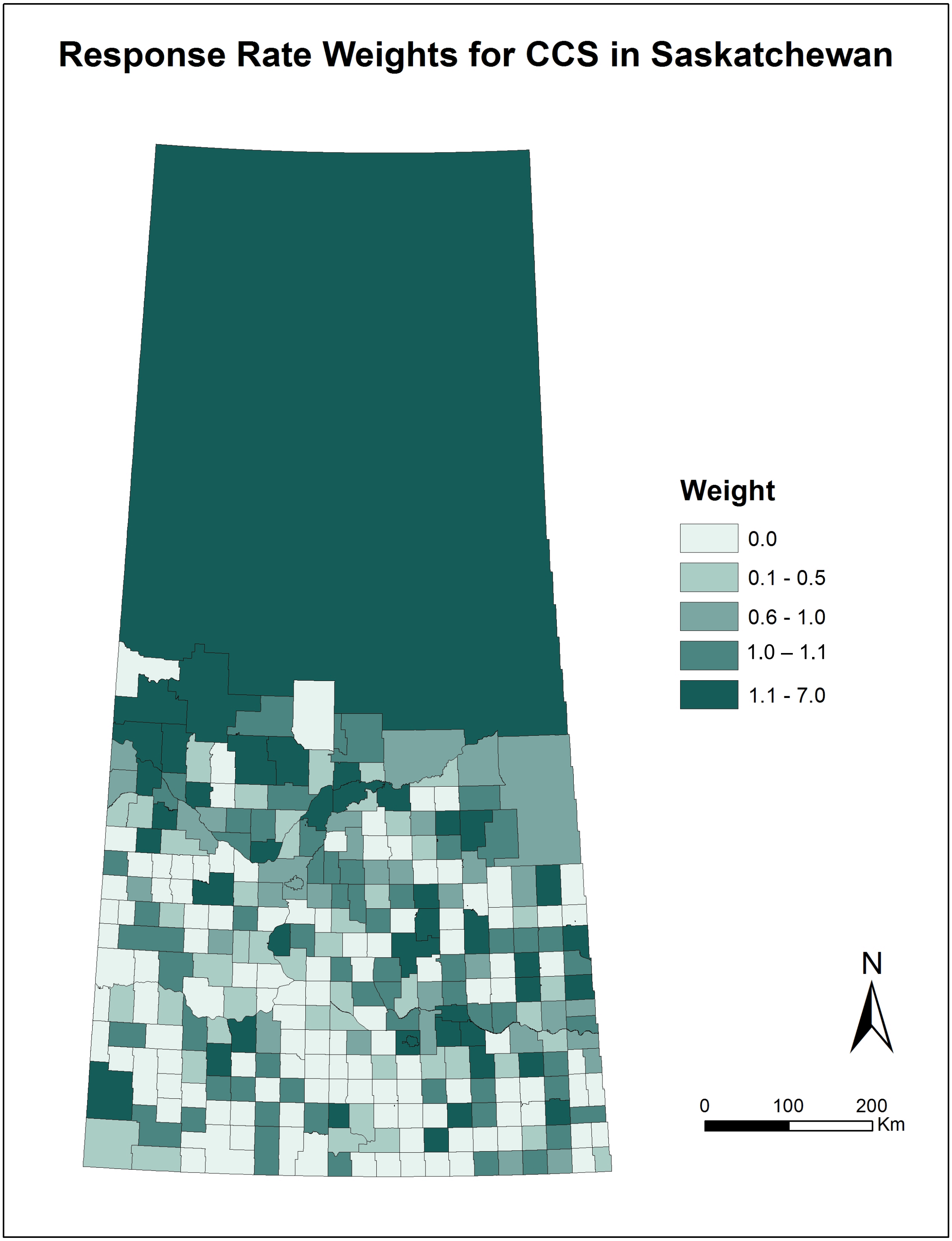

Figure 1: Response rate weights for each CCS in Saskatchewan (zero [light green] – no completed survey in CCS; darker shades of green – representative sample for CCS (if weight is exactly 1) or undersampled CCS; three other quintiles – oversampled CCS)

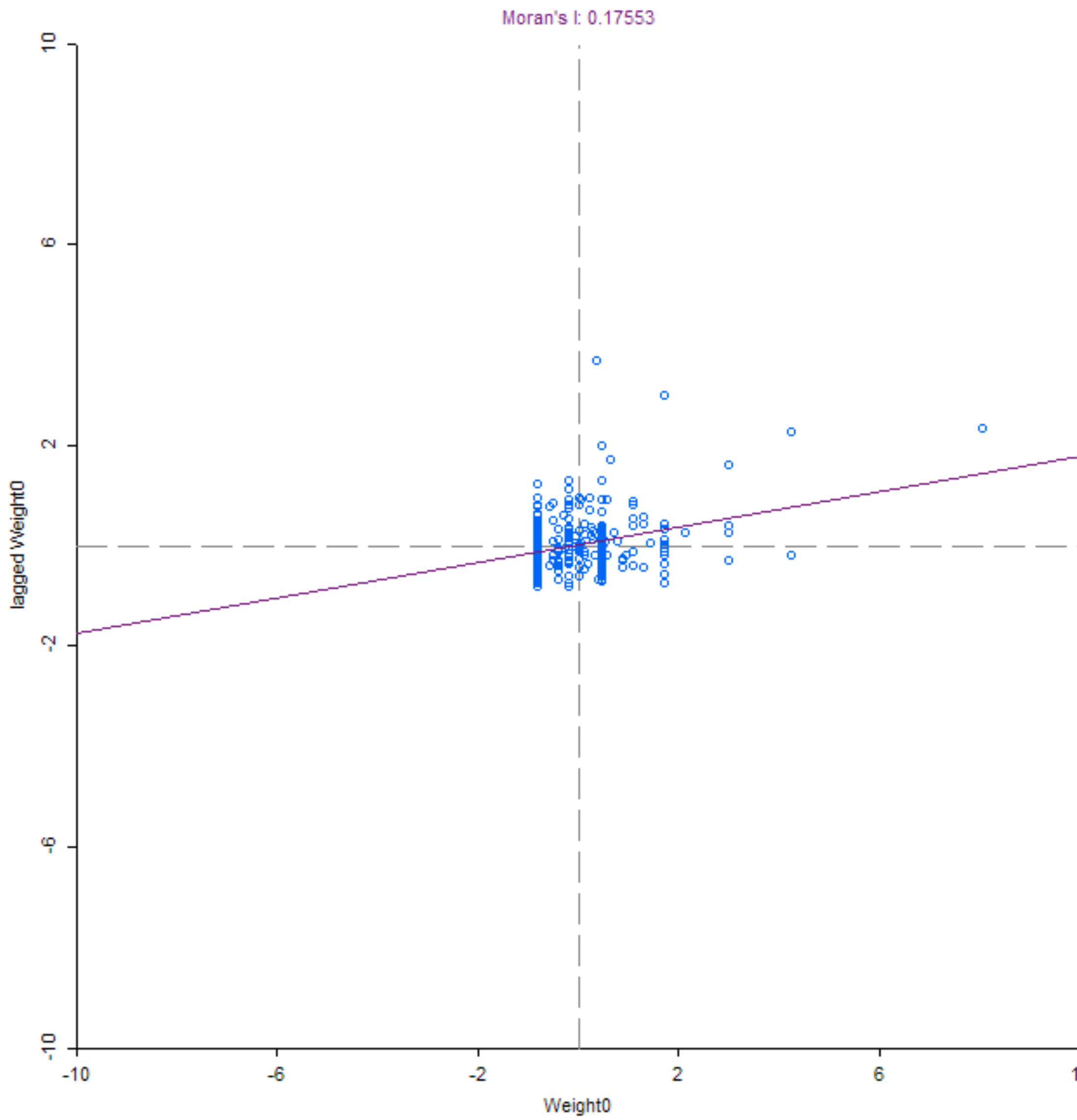

Figure 2: Univariate Moran's I for variable Weight

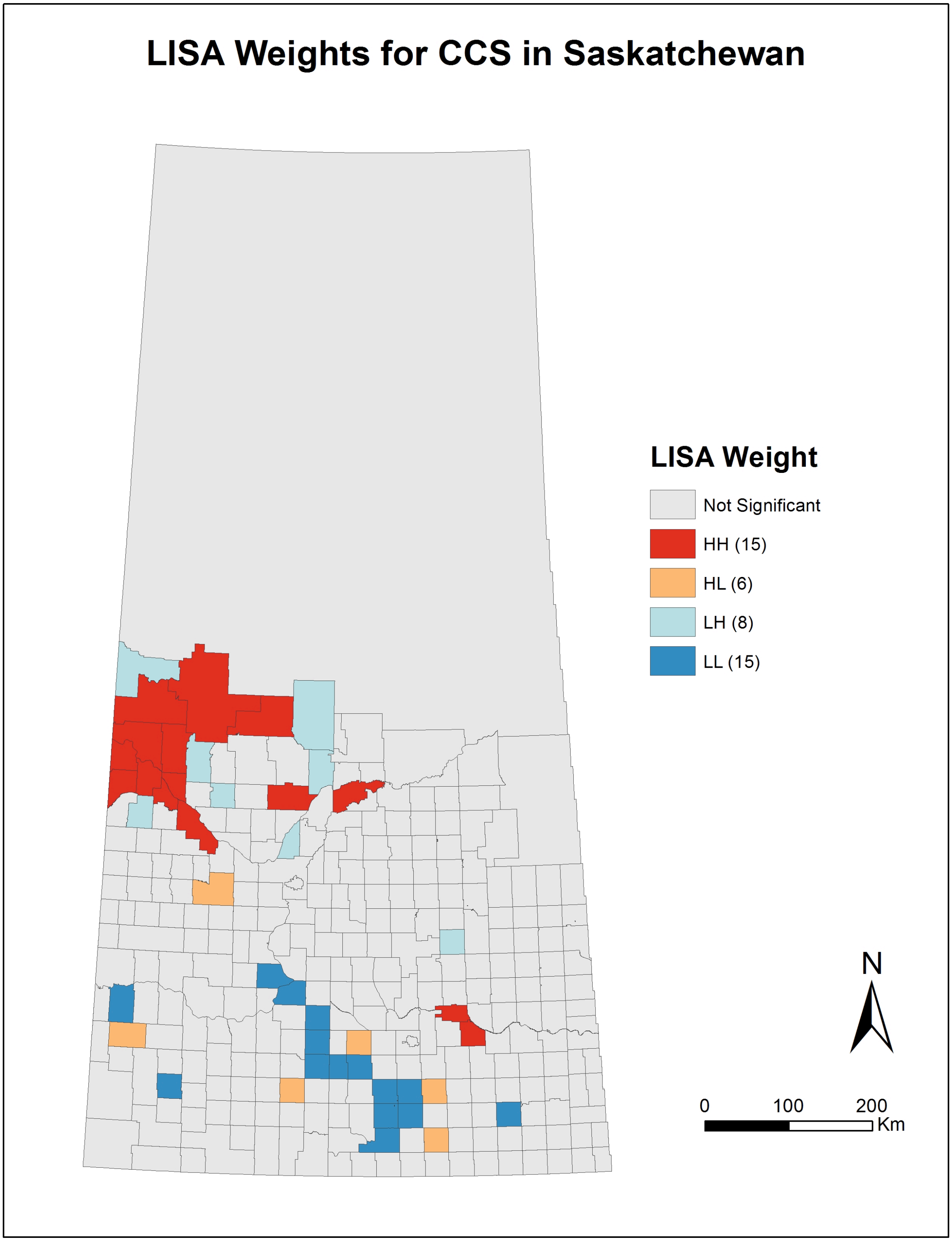

Figure 3: LISA of weights for CCS in Saskatchewan

Conclusions

Overall, considering Moran's I, spatial autocorrelation in the sampling error (weight) does not seem to be a major problem in our data set. However, there is a large number of CCS (a third) for which there were no completed surveys (with expected survey numbers ranging from 0 to 3) and hence the weight could not be calculated. In addition, there is a noticeable cluster of undersampled CCS in Northwestern Saskatchewan (North of Lloydminster). This cluster may have occurred by chance, but might also be an indicator of particularly low response rates or a problem in our sample strategy. It would be interesting to run the same analysis for other Saskatchewan-wide surveys conducted at the Social Sciences Research Laboratories. In addition, if there is interest in conducting a statistical analysis at the level of the consolidated census subdivisions, then the overall sample size needs to be increased. With the size of our current data set, an analysis based on census divisions (n=18) or a spatial division into five regions (Saskatoon CMA, Regina CMA, Smaller Cities, Rural South, Rural North) is more appropriate.

References

Anselin, L. (1995). Local indicators of spatial association – LISA. Geographical Analysis, 27, 93–115.

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. Geographical Analysis, 38, 5–22.

Bell S. 2009. Experimental Design. In Kitchin R, Thrift N (eds) International Encyclopedia of Human Geography,Volume 3, pp. 672–675. Oxford: Elsevier.

Bell, S., Jones, M., & Wei, T. (2013). *Communicating with uncertain data: A search for spatial and geographic patterns* Paper presented at the COSIT 2013 Workshop on Visually-Supported Reasoning with Uncertainty, Scarborough, North Yorkshire, UK.

Bell, S., & Wei, T. (2014). Mapping the Spatial Pattern of the Uncertain Data: A comparison of Global Non-response Rate (GNR) between Metropolitan and Non-Metropolitan Area. *Spatial Knowledge and Information - Canada*.

Berdahl, L., Bell, S., Bourassa, M., & Jana, F. (2014). Trust, Policymaking, and the Nuclear Sector: Overview of Saskatchewan Attitudes.

Bourassa, M., Bell, S., Berdahl, L., & Jana, F. (2014). Values, Knowledge, and the Nuclear Sector: Overview of Saskatchewan Attitudes.

Fried, J., Bell, S., Berdahl, L., & Bourassa, M. (2014a). The Nuclear Energy Sector: Overview of Saskatchewan Attitudes.

Fried, J., Bell, S., Berdahl, L., & Bourassa, M. (2014b). The Nuclear Sector: Nuclear Medicine, Uranium Mining, and Nuclear Fuel Waste Management. Overview of Saskatchewan Attitudes.

Shah, T. I., Aspen, L., & Bell, S. (2014). Assessment of Choice of Units of Analysis for Studying Associations between Geographic Accessibility to PHC Services and Socio-demographic Factors. *Spatial Knowledge and Information - Canada, 2*.

Vivier, P. M., Hauptman, M., Weitzen, S. H., Bell, S., Quilliam, D. N., & Logan, J. R. (2011). The important health impact of where a child lives: Neighborhood characteristics and the burden of lead poisoning. *Maternal and Child Health Journal, 15*(8), 1195-1202.

Singh, K. (2007). Quantitative Social Research Methods. New Delhi: Sage Publications.
Statistics Canada (2013). Census consolidated subdivisions (CCS). Retrieved from http://www.statcan.gc.ca/pub/92-195-x/2011001/geo/ccs-sru/ccs-sru-eng.htm