Robert A. Yaffee, Ph.D. NSF report

5 October 2009

Research, Writing, and teaching related to the Chernobyl project October 1, 2008-July 5, 2009.

Research Design and sampling for grant October – November 2008. Random digit dialing design of sampling protocol. Generation of randomly selected numbers constrained by restrictions on the first few numbers of different raions (administrative subdivision similar to a U. S. county). I randomly generated phone numbers proportional to the number of phones for the raion (Ukrainian county). A list of the number of phone numbers in each raion was provided by the telephone company at some cost. We travelled to the Kiev, Ukraine on 1 October 2008, where colleagues assisted in recruiting teams of interviewers. The interviewers were trained, in which I explained our need for geographical coordinates to identify the history of residence, work, and vacation of our respondents since the Chernobyl disaster... We needed a common format for recording the latitude and longitude measures we would use to do our computerized mapping of radiation diffusion, physical and mental illness history, to sequentially test the spatial correlation between our radiation dose reconstruction and spatial patterns of physical and psychological illness over several points in time. I had written a computer program to convert degrees, minutes, and seconds to degrees in decimal form and taught that to our coordinators in Kiev. Eventually, we decided to use Google earth for finding the coordinates of reported locations because hard-copy listings of map coordinates were not generally available in the Ukraine. During the Soviet era, only the military had this data. After finishing my lecture and training, I returned to the United States.

The research design began while I was in Kiev. I began the sampling frame construction and set up the number of phone numbers estimated to be needed for recruitment of respondents. We estimated a response rate of 1/8 to those who were called. Therefore, I generated 8 times the number of phone numbers we needed for our pilot study, in such a manner that the proportions of numbers generated in each raion were proportional to the number of phone listed by the telephone company in each raion on the phone tallies per raion that they supplied. Using the randomly generated phone numbers, people were contacted and asked for permission to interview. When people did not answer, they were called back four times. After obtaining acceptances, appointments made, and interviews for the pilot have been completed. The data has been scanned in. We finished the first phase of cleaning the data of scanning errors. We are in the second phase of data cleaning, where we conduct the range and consistency checks. We are seeing which questions are working and which are not. We are translating Russian to English labels for our value labels and translating ICD9 and ICD10 codes for construction of our value labels. We are reviewing frequency-percentage tabulations to be sure that there no scores outside the permitted ranges. We will soon be doing the consistency checks of correlations between items to be sure of the coding. As we clean the pilot data, we provide regular feedback to our interview coordinators to be sure that any problems are prevented or resolved.

Preliminary spatial orientation: To give us perspective and focus, we sought maps of the radiation pollution in the Ukraine. We managed to get MapInfo data from someone who had done some work in this area. The map revealed CS137 pollution contours around various locations in the Ukraine. I immediately accessed the data and reconstructed the MapInfo map of this pollution and tested various levels of zoom before distributed several of them to my colleagues. I converted the MapInfo data to an ArcGIS dataset, with which I was able to generate multiple maps of finer resolution, which I promptly distributed them to my colleagues. Upon reviewing the maps, we came up with a set of follow-up questions we had for our map vendor. These maps were generated and distributed largely before February of this year.

Spatial Analysis Preparation. I realized that there is much preparation for a space time analysis of radiation pollution of a specific isotope-- CS137. We needed polygon shape files that defined the boundaries of the Ukrainian Oblasts (states) and raions (counties). We needed these in order to geographically map the radiation contamination at different times and to represent at several points in time the geographical distribution of physical and mental illness we found. While in Kiev, we made contact with someone who had done mapping of this sort of thing before and managed to obtain from him maps of the Ukraine and former counts of radiation pollution. These maps were MapIinfo maps. We want to store our dataset in SAS files owing to the excellent data management power and flexibility of SAS. Although we plan to do our mapping in ArcGis, we want to do our spatial analysis in S-Plus or, R. I later decided that for matters of this complexity, Geobugs might be a good alternative approach. I managed to obtain the boundary coordinates from the MapInfo map and convert them to files that could be read by S-Plus and R. I also managed to generate a polygon shape file that would provide the geographical frame of reference for a Bayesian disease mapping as well. Meanwhile, I extracted the latitude and longitude coordinates for the radiation contours of the map provided to us. We are in the process now of determining the time interval over which these counts were measured. From the decay rates of the isotopes, we should be able to compute the levels of radiation of that isotope at the measured location at any point in time. This will provide us with an external dose for all those who lived in that area. Once the survey is completed, we hope that there will be enough information about the whereabouts of the respondent and his diet to reconstruct his accumulated dosage since the Chernobyl catastrophe. This assessment will then a factor in our estimate of possible candidate sources of illness, either physical, mental, or both. Thus, the preparation of polygon shape files for spatial analysis in S-Plus, Winbugs, and Geobugs has been completed.

Preliminary preparation. On Dec 10-13, 2008 I took a Bayesian disease mapping course at Medical University of South Carolina, where we studied Poisson models with random effects, conditional autoregressive (CAR) models, spatial logistic model s for threshold estimation, space-time models count models with possible interactions for space and time. The purpose was to explore alternative approaches to spatial analysis.

On May 10-11, 2009 I followed this up with an Advanced Bayesian Disease mapping course in Medical University of South Carolina to explore more alternatives in preparation for our analysis. In this course, we studied hierarchical Bayesian models for disease mapping. We studied model critiques-dealing with DIC evaluation of models, spatial trends and spatially aligned covariates, convolution models with both correlated and uncorrelated heterogeneity( ch and uh, respectively), proper CAR models, full multivariate normal models, mixture models, as well as Bayesian kriging for mapping, conditional logistic spatial models for threshold effects, use of Delauny triangulation to generate Dirichelet tessellation to designate neighbors for an adjacency matrix, hierarchical modeling and testing of sequentially added effects from distance(D) only, D + uh, D+uh+age(A), and finally the D+A,D+A+uh+ch model. We need to compare models with the DIC. We examined ZIP and ZIB models, sparse Poisson convolution models, geographically weighted random coefficient models with spatial correlation, mixture convolution models for modeling spatial discontinuities. In addition we examined spatial survival models with censoring issues along with imputation for such scores, ecological hierarchical models with random contextual effects, latent variable models, multivariate models for comorbidity (for example, asthma and COPD), competing risk models, common component dynamic models, spatial retrospective models for rare diseases, interactions and their prior distributions, and the application of the Kalman filter for space-time models with Poisson measurement models and autoregressive transition models, and the Hidden Markov models. We addressed biosurveillance in space-time models with a focus on exceedance probabilities or changes in variance to identify hot-spots or warnings of emerging disease clusters.

My teaching with respect to this project has included a lecture on an Introduction to Bayesian Disease Mapping, Academy of Labor and Social Relations in Kiev, Ukraine on October 6, 2008**.**